

Editorial Board

Simone Diniz Junqueira Barbosa

*Pontifical Catholic University of Rio de Janeiro (PUC-Rio),
Rio de Janeiro, Brazil*

Phoebe Chen

La Trobe University, Melbourne, Australia

Alfredo Cuzzocrea

ICAR-CNR and University of Calabria, Italy

Xiaoyong Du

Renmin University of China, Beijing, China

Joaquim Filipe

Polytechnic Institute of Setúbal, Portugal

Orhun Kara

TÜBİTAK BİLGEM and Middle East Technical University, Turkey

Tai-hoon Kim

Konkuk University, Chung-ju, Chungbuk, Korea

Igor Kotenko

*St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences, Russia*

Dominik Ślęzak

University of Warsaw and Infobright, Poland

Xiaokang Yang

Shanghai Jiao Tong University, China

Chunfeng Liu Leizhen Wang
Aimin Yang (Eds.)

Information Computing and Applications

Third International Conference, ICICA 2012
Chengde, China, September 14-16, 2012
Proceedings, Part I

Volume Editors

Chunfeng Liu
Hebei United University
College of Sciences
Tangshan, Hebei, China,
E-mail: chunfengliu201@163.com

Leizhen Wang
Northeastern University
Qinhuangdao, Hebei, China,
E-mail: wangleizhen201@163.com

Aimin Yang
Hebei United University
College of Sciences
Tangshan, Hebei, China,
E-mail: yangaimin201@163.com

ISSN 1865-0929

ISBN 978-3-642-34037-6

DOI 10.1007/978-3-642-34038-3

Springer Heidelberg Dordrecht London New York

e-ISSN 1865-0937

e-ISBN 978-3-642-34038-3

Library of Congress Control Number: 2012948331

CR Subject Classification (1998): F.1.2, H.3.4, H.3.5, G.3, H.2.7, H.2.8, K.6, C.2.1, C.2.4, J.1, J.3, J.7

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Welcome to the proceedings of the 3rd International Conference on Information Computing and Applications (ICICA 2012), which was held during September 14–16, 2012, in Mountain Resort, Chengde, China.

As future generation information technology, information computing, and applications become more and more specialized, information computing and applications including hardware, software, communications, and networks are growing with ever increasing scale and heterogeneity, and becoming overly complex. The complexity is becoming more and more critical as the number of applications increase. To cope with the growth and computing complexity, information computing and applications are focusing on intelligent, selfmanageable, scalable computing systems and applications to the maximum extent possible without human intervention or guidance.

With the rapid development of information science and technology, information computing has become the third approach of science research. Information computing and applications is the field of study concerned with constructing intelligent computing, mathematical models, and numerical solution techniques and using computers to analyze and solve natural scientific, social scientific, and engineering problems. In practical use, it is typically the application of computer simulation, intelligent computing, internet computing, pervasive computing, scalable computing, trusted computing, autonomy-oriented computing, evolutionary computing, mobile computing, applications, and other forms of computation to problems in various scientific disciplines and engineering. Information computing and applications is an important underpinning for techniques used in information and computational science and there are many unresolved problems worth studying.

The ICICA 2012 conference provided a forum for engineers and scientists from academia, industry, and government to address the most innovative research and development including technical challenges and social, legal, political, and economic issues, and to present and discuss their ideas, results, work in progress, and experience on all aspects of information computing and applications.

There was a very large number of paper submissions (1089). All submissions were reviewed by at least three Program or Technical Committee members or external reviewers. It was extremely difficult to select the presentations for the conference because there were so many excellent and interesting submissions. In order to allocate as many papers as possible and keep the high quality of the conference, we finally decided to accept 330 papers for presentation, reflecting a 30.3% acceptance rate. A total of 115 papers have been included in this volume. We believe that all of these papers and topics not only provide novel ideas, new results, work in progress, and state-of-the-art techniques in this field, but will

also stimulate future research activities in the area of information computing and applications.

The exciting program for this conference was the result of the hard and excellent work of many others, such as Program and Technical Committee members, external reviewers, and Publication Chairs under a very tight schedule. We are also grateful to the members of the Local Organizing Committee for supporting us in handling so many organizational tasks, and to the keynote speakers for accepting to come to the conference with enthusiasm. Last but not least, we hope you enjoyed the conference program, and the beautiful attractions of Chengde, China.

September 2012

Yanchun Zhang
Baoxiang Liu
Chunfeng Liu

Organization

ICICA 2012 was organized by Hebei United University, Hebei Scene Statistical Society, and sponsored by the National Science Foundation of China, Hunan Institute of Engineering, Yanshan University, Northeastern University at Qinhuangdao and Chengde Petroleum College. It was held in cooperation with *Communications in Computer and Information Science* (CCIS) of Springer.

Executive Committee

Honourary Chair

Jun Li Hebei Polytechnic University, China

General Chairs

Yanchun Zhang University of Victoria, Australia
Baoxiang Liu Hebei Polytechnic University, China

Program Chairs

Chunfeng Liu Hebei Polytechnic University, China
Fengbo Hou Chengde Petroleum College, China
Wenjiang Du Chongqing Normal University, China

Local Arrangements Chairs

Jincai Chang Hebei Polytechnic University, China
Aimin Yang Hebei Polytechnic University, China

Steering Committee

Qun Lin	Chinese Academy of Sciences, China
Maode Ma	Nanyang Technological University, Singapore
Nadia Nedjah	State University of Rio de Janeiro, Brazil
Lorna Uden	Staffordshire University, UK
Yiming Chen	Yanshan University, China
Changcun Li	Hebei Polytechnic University, China
Zhijiang Wang	Hebei Polytechnic University, China
Rongbo Zhu	South-Central University for Nationalities, China
Guohuan Lou	Hebei Polytechnic University, China
Jixian Xiao	Hebei Polytechnic University, China
Xinghuo Wan	Hebei Polytechnic University, China

Chunying Zhang
Dianchuan Jin

Hebei Polytechnic University, China
Hebei Polytechnic University, China

Publicity Chairs

Aimin Yang
Xilong Qu

Hebei Polytechnic University, China
Hunan Institute of Engineering, China

Publication Chairs

Yuhang Yang

Shanghai Jiao Tong University, China

Financial Chair

Jincai Chang

Hebei Polytechnic University, China

Local Arrangements Committee

Lihong Li
Shaohong Yan
Yamian Peng
Lichao Feng
Yuhuan Cui

Hebei Polytechnic University, China
Hebei Polytechnic University, China
Hebei Polytechnic University, China
Hebei Polytechnic University, China
Hebei Polytechnic University, China

Secretaries

Kaili Wang
Jingguo Qu
Yafeng Yang

Hebei Polytechnic University, China
Hebei Polytechnic University, China
Hebei Polytechnic University, China

Program/Technical Committee

Yuan Lin

Yajun Li
Yanliang Jin
Mingyi Gao
Yajun Guo
Haibing Yin
Jianxin Chen
Miche Rossi
Ven Prasad

Norwegian University of Science and
Technology, Norway
Shanghai Jiao Tong University, China
Shanghai University, China
National Institute of AIST, Japan
Huazhong Normal University, China
Peking University, China
University of Vigo, Spain
University of Padua, Italy
Delft University of Technology,
The Netherlands

Mina Gui
Nils Asc
Ragip Kur
On Altintas

Texas State University, USA
University of Bonn, Germany
Nokia Research, USA
Toyota InfoTechnology Center, Japan

Suresh Subra	George Washington University, USA
Xiyin Wang	Hebei Polytechnic University, China
Dianxuan Gong	Hebei Polytechnic University, China
Chunxiao Yu	Yanshan University, China
Yanbin Sun	Beijing University of Posts and Telecommunications, China
Guofu Gui	CMC Corporation, China
Haiyong Bao	NTT Co., Ltd., Japan
Xiwen Hu	Wuhan University of Technology, China
Mengze Liao	Cisco China R&D Center, China
Yangwen Zou	Apple China Co., Ltd., China
Liang Zhou	ENSTA-ParisTech, France
Zhanguo Wei	Beijing Forestry University, China
Hao Chen	Hu'nan University, China
Lilei Wang	Beijing University of Posts and Telecommunications, China
Xilong Qu	Hunan Institute of Engineering, China
Duolin Liu	ShenYang Ligong University, China
Xiaozhu Liu	Wuhan University, China
Yanbing Sun	Beijing University of Posts and Telecommunications, China
Yiming Chen	Yanshan University, China
Hui Wang	University of Evry in France, France
Shuang Cong	University of Science and technology of China, China
Haining Wang	College of William and Mary, USA
Zengqiang Chen	Nankai University, China
Dumisa Wellington Ngwenya	Illinois State University, USA
Hu Changhua	Xi'an Research Inst. of Hi-Tech, China
Juntao Fei	Hohai University, China
Zhao-Hui Jiang	Hiroshima Institute of Technology, Japan
Michael Watts	Lincoln University, New Zealand
Tai-hon Kim	Defense Security Command, Korea
Muhammad Khan	Southwest Jiaotong University, China
Seong Kong	The University of Tennessee, USA
Worap Kreesuradej	King Mongkuts Institute of Technology Ladkrabang, Thailand
Uwe Kuger	Queen's University of Belfast, UK
Xiao Li	CINVESTAV-IPN, Mexico
Stefanie Lindstaedt	Know-Center, Austria
Paolo Li	Polytechnic of Bari, Italy
Tashi Kuremoto	Yamaguchi University, Japan
Chun Lee	Howon University, South Korea
Zheng Liu	Nagasaki Institute of Applied Science, Japan
Michiharu Kurume	National College of Technology, Japan

Sean McLoo	National University of Ireland, Ireland
R. McMenemy	Queen's University Belfast, UK
Xiang Mei	The University of Leeds, UK
Cheol Moon	Gwangju University, South Korea
Veli Mumcu	Technical University of Yildiz, Turkey
Nin Pang	Auckland University of Technology, New Zealand
Jian-Xin Peng	Queen's University of Belfast, UK
Lui Piroddi	Technical University of Milan, Italy
Girij Prasad	University of Ulster, UK
Cent Leung	Victoria University of Technology, Australia
Jams Li	University of Birmingham, UK
Liang Li	University of Sheffield, UK
Hai Qi	University of Tennessee, USA
Wi Richert	University of Paderborn, Germany
Meh shafiei	Dalhousie University, Canada
Sa Sharma	University of Plymouth, UK
Dong Yue	Huazhong University of Science and Technology, China
YongSheng Ding	Donghua University, China
Yuezhi Zhou	Tsinghua University, China
Yongning Tang	Illinois State University, USA
Jun Cai	University of Manitoba, Canada
Sunil Maharaj Sentech	University of Pretoria, South Africa
Mei Yu	Simula Research Laboratory, Norway
Gui-Rong Xue	Shanghai Jiao Tong University, China
Zhichun Li	Northwestern University, China
Lisong Xu	University of Nebraska-Lincoln, USA
Wang Bin	Chinese Academy of Sciences, China
Yan Zhang	Simula Research Laboratory and University of Oslo, Norway
Ruichun Tang	Ocean University of China, China
Wenbin Jiang	Huazhong University of Science and Technology, China
Xingang Zhang	Nanyang Normal University, China
Qishi Wu	University of Memphis, USA
Jalel Ben-Othman	University of Versailles, France

Table of Contents – Part I

Internet Computing and Applications

WIPOMTS: An Internet Public Opinion Monitoring System	1
<i>Shouhua Zhang, Kaiyu Shang, Shuai Cong, Bin Zhang, and Zhenpeng Liu</i>	
Social Context Enabled Description Model for Web Services	9
<i>Xiao Zheng, Qingyong Wu, Dongliang Ke, Haixia Li, and Yaqing Shi</i>	
Improved Learning Algorithm for Self-adaptive Neural Nets Based on Principal Component Analysis	16
<i>Shujie Du</i>	
Digital Library Network Based on the San Technology	25
<i>Jiao Cui-Ling and Wang Jian-ping</i>	
Websites Based on J2EE of Struts	31
<i>Jiao Cui-Ling and Wang Jian-ping</i>	
Research of a Vertical Search Engine for Campus Network	37
<i>Rujia Gao, Wanlong Li, Shanhong Zheng, and Hang Li</i>	
Efficient Control Scheme for Surface Temperature of Hot Roller Based on Neural Network	44
<i>XueLi Wu, Lifeng Hou, and Hui An</i>	
Online Shopping System Based on Improved MD5	51
<i>Ji-Rui Li and Su-Qing Zhang</i>	
Adaptivity in Location-Based Services	58
<i>Rui Zhou, Wensheng Guo, and Nan Sang</i>	
An Evolution Model of Emotional Internet Public Opinion with Informed Marks	66
<i>Wu Yu, Hao Wang, Xian Wang, and Jianbo Li</i>	
Research of LSH and Outliers Detection	74
<i>Ying-yan Wang, Rui Zeng, Ming-zhong Li, and Fang Li</i>	
Research on P2P Botnet Network Behaviors and Modeling	82
<i>Huabo Li, Guyu Hu, and Yun Yang</i>	
Web-Based Solution for the Diversity of Equipment	90
<i>Chang Yan-shuo and Hua Qing-yi</i>	

Fusion Assessment Methods for Bridge Health State Based on Two Step Neural Networks Ensemble	97
<i>Li Shan, Zhao Hui, and Rao Hong</i>	

Multimedia Networking and Computing

Classification of Rice Appearance Quality Based on LS-SVM Using Machine Vision	104
<i>Xiai Chen, Shuang Ke, Ling Wang, Hong Xu, and Wenquan Chen</i>	
Research of University Disaster Recovery System Based on Virtualization Technology	110
<i>Yue Li, Jian Sun, Zheng Zheng, and Hong Shen</i>	
Research of Self-study English Pronunciation on WindowsCE	117
<i>Shen Hong, Li Yue, and Wang Jinfang</i>	
Study of Linearization for Spectral Imaging Device	125
<i>Songhua He</i>	
Silverlight-Based Distance Teaching Application	133
<i>Jin Wang</i>	
Uncorrelated Discriminant Isometric Projection for Face Recognition ...	138
<i>Bo Ge, Yanling Shao, and Yunxing Shu</i>	
Design and Realization of Multimedia Animation Based on Flash	146
<i>Shouping Wang and Zhenbing Wang</i>	
The Ship Propeller Design System for Coupling of CAD, CAE and CAO	152
<i>Jinfeng Huang</i>	

Skeleton Extraction of a Specified Object in the Gray Image Based on Geometric Features	161
<i>Zhihui Yang, Fangfang Guo, and Ping Dong</i>	

Detection and Correction Scheme of Internet Chat Lingo Based on Statistic and Pinyin Similarity	169
<i>Bo Han and Zhengwen Li</i>	

Intelligent Computing and Applications

A Filling Algorithm of Mining Constraint Frequent Neighboring Class Set	175
<i>Ling Hu, Yao Li, Jiang Xiong, and Gang Fang</i>	

Sea Clutter Constant False-Alarm Processing Technology Research Based on Wavelet Transform	183
<i>Shuang Xu, Yanqiu Cui, and Jifeng Ding</i>	
A Tableau-Based Reasoning Algorithm for Distributed Dynamic Description Logics	192
<i>Zhuxiao Wang, Zhitao Guan, Wei Li, Kehe Wu, Jing Guo, and Guanhua Tian</i>	
Improvement of Target Extraction and Dense Matching	200
<i>Zetao Jiang, Qiang Wang, and Yanru Cui</i>	
Intelligent Scheme of Removing Plug Oil Well Based on Mobile Robot	208
<i>Xiaojie Tian, Yonghong Liu, Yulong Xu, Rongju Lin, Wei Deng, and Rui Zhang</i>	
A Novel Meta-heuristic for the Multi-depot Vehicle Routing Problem ...	216
<i>Jianping Luo, Xia Li, and Min-Rong Chen</i>	
An Improved Association Rule Privacy Protection Algorithm Research	225
<i>Yiwu Chen, Qingshui Li, and Lingna He</i>	
Multi-step Prediction of Volterra Neural Network for Traffic Flow Based on Chaos Algorithm	232
<i>Lisheng Yin, Yigang He, Xueping Dong, and Zhaoquan Lu</i>	
Improved Calculation Scheme of Structure Matrix of Boolean Network Using Semi-tensor Product	242
<i>Jinyu Zhan, Shan Lu, and Guowu Yang</i>	
Adaptive RBF Neural Network Filtering Predictive Model Based on Chaotic Algorithm	249
<i>Lisheng Yin, Yigang He, Xueping Dong, and Zhaoquan Lu</i>	
Estimation of Leaf Weight Based on Snowflake Theory	258
<i>Lihui Zhou and Hong Wang</i>	
Efficient Decision-Making Scheme Based on LIOWAD	265
<i>Meihua Cheng and Shouzhen Zeng</i>	
Computational Statistics and Applications	
Analysis of Sensitive Questions of MSM Based on RRT	273
<i>Wei Li, Ge Gao, Yu-hua Ruan, Xiang-yu Chen, and Qiao-qiao Du</i>	
Cointegration Analysis Based on Linear Error in Variable Model	280
<i>Mengtian Yue and Huirong Cao</i>	

Statistical Analysis of Polycyclic Aromatic Hydrocarbons Projects	288
<i>Jianguo Feng, Guihua Li, Jintang Zhang, and Le Xia</i>	
Research on CNC Contour Error Compensation Approach Based on Straightway Approximate Nodes	294
<i>Guoyong Zhao, Zhiyong Li, and Hongjing An</i>	
Least Squares η – Hermitian Solution for Quaternion Matrix Equation $AXB = C$	300
<i>Shifang Yuan</i>	
Reasonable Arrangements for Ophthalmic Beds Based on Linear Programming	306
<i>Lihui Zhou, Shaohong Yan, and Yongsheng Zhang</i>	
Error Analysis of Generalized LxF Schemes for Linear Advection Equation with Damping	313
<i>Yun Wu and Dong Qu</i>	
Generalized Lax-Friedrichs Scheme for Convective-Diffusion Equation	321
<i>Wei Tong and Yun Wu</i>	
Fuzzy Time Series Forecasting Based on Weber-Fischna Law	329
<i>Lihong Li, Yan Gao, and Yan Wang</i>	
Oscillation of Runge-Kutta Methods for a Scalar Differential Equation with One Delay	336
<i>Qi Wang, Jiechang Wen, and Fenglian Fu</i>	
Analysis of Wall Criteria in Numerical Computation of Shock Wave/Layer Interaction	345
<i>Xinwen Xu and Yong Yu</i>	
A Voxelization Algorithm for 3D Body-Centered Cubic Line Based on Adjunct Parallelepiped Space	352
<i>Lijun He, Yongkui Liu, Degao Wang, and Jian Yun</i>	
Jackson's Theorem in Hardy-Sobolev Type Spaces in the Unit Polydiscs	360
<i>Yingwei Chen, Zhijun Wang, and Wenlei Dong</i>	
On the Upper Bounds of Local Inverse Signed Edge Domination Numbers in Graphs	368
<i>Zhifang Feng, Wensheng Li, Huaming Xing, and Qiongyu Ma</i>	
Forecast of Employment Based on Independent Component Analysis	373
<i>Lu Tan and Han Zhang</i>	
Complete Convergence for NA Random Sequence	382
<i>Yourong Wang, Yanli Liu, and Yili Tan</i>	

A Remark on Distance Regular Graph with $k=10$, $a_1=1$	387
<i>Junhong Ma and Bing Han</i>	

Knowledge Management and Applications

Information Entropy, Similarity Measure and Inclusion Measure of Intuitionistic Fuzzy Sets	392
<i>Qiansheng Zhang, Fuchun Liu, Lihua Wu, and Shihua Luo</i>	
Optimization of Tandem Cold Rolling Schedule Based on Collaborative Optimized PSO	399
<i>Li Ying, Wang Jing-sheng, Wang Hong-rui, and Wei Li-xin</i>	
Career Decision-Making of University Students Based on Analytic Hierarchy Process	408
<i>Ming Lu</i>	
Study on Information Literacy of College Students with University Library	414
<i>Hai-bo Yang, Yan-yan Zhang, and Ning Yu</i>	
Comparison of Different Expanded GED-GARCH for Exchange Rate Volatility	421
<i>Yan Gao, Dongmei Li, and Lihui Zhou</i>	
Component Description Model Based on Software Architecture	427
<i>Zhiyu Chen, Liufu Song, Junqiang Liu, Ming Hu, and Rujia Gao</i>	
Electronic Transaction Scheme Based on Fast ECC Multiplication	433
<i>Shouzhi Xu, Bo Xu, Pengfei Guo, and Chenxia Li</i>	
Fine Granularity Pipeline Indexing Algorithm for Multi-core Platform	441
<i>Su-Qing Zhang and Ji-Rui Li</i>	
Collaborative Filtering Algorithm Based on the Preference List in the Digital Library	448
<i>Gao Feng, Liu Zhiyou, and Liu Huixin</i>	
Customer Threshold Strategies in Observable Queues with Partial Information of Service Time	456
<i>Wei Sun and Shiyong Li</i>	
Fault Diagnosis Based on Low-Frequency Acceleration Signals towards Large-Scale Recreational Facilities	463
<i>Gen Li, Yu-tian Zhu, and Shuan-zhong Wang</i>	

Communication Technology and Applications

Research on Embedded-Based Wireless Authentication System	469
<i>Wei Pan, Xiang Yang, Yuanyi Zhang, and Zhe Guo</i>	
Data Reliability Analysis for Flooding Transmission of Wireless Sensor Networks	476
<i>Cao Yuan-Yuan and Feng Hai-Lin</i>	
Effective Management and Utilization of Mobile Learning Based on Wireless Campus Network	484
<i>Shujie Du, Zhuan Zhou, Lanfang Liu, Ling Jiao, and Bo Wu</i>	
Simulation of Malicious Nodes Detection Based on Machine Learing for WSN	492
<i>Kuancheng Zou, Yuanling Ouyang, Chuncheng Niu, and Yi Zou</i>	
A Robotic System Based on Wireless Control.....	500
<i>Changjiu Pu</i>	
Study on Wireless Campus Mesh Routing Protocol Based on Security Association	507
<i>Limei Fan</i>	
Mobile Counteracted Time Synchronization for Underwater Sensor Networks	515
<i>Ying Guo and Yutao Liu</i>	
Cross-Layer Routing Metric for Wireless Mesh Networks	523
<i>Xiang Li, Yong-Jian Yang, Chun-Sheng Cui, and Guan-Nan Yang</i>	
Optimization and Realization of Ad Hoc On-Demand Distance Vector Routing Based on NS2.....	531
<i>Deyu Lin and Xiaojing Meng</i>	
A Double-Links Scheme for DSDV in Mobile Ad Hoc Networks	538
<i>Jia Lu, Junhui Wang, Gang Han, Baoliang Li, and Wenhua Dou</i>	
Time Frequency Analysis of Underwater Carrier-Free Pulse.....	548
<i>Yunlu Ni and Hang Chen</i>	
Data Aggregation Scheme Based on Compressed Sensing in Wireless Sensor Network	556
<i>Guangsong Yang, Mingbo Xiao, and Shuqin Zhang</i>	
Study of Dynamic Traffic Flow Network Model Based on LWR Model	562
<i>Wang Jingbo, Qian Li, and Wang Jingtao</i>	

Multiple Sources-Multiple Relays Precoding Algorithm in Cooperative Wireless Networks	569
<i>Yongli An, Xuefei Gao, and Kaiyu Zhu</i>	
Efficient Detect Scheme of Botnet Command and Control Communication	576
<i>Jinquan Zeng, Weiwen Tang, Caiming Liu, Jianbin Hu, and Lingxi Peng</i>	
H.264 Decoder Based on Motion Compensation Optimizer	582
<i>Xiaohong Hao and Jiaqi Li</i>	

Influence of Scattering on Laser Communication	590
<i>Junjun Wu and Mingchuan Meng</i>	

Research on University Library Services in Science and Technology Novelty Search	597
<i>Lei Mei-Shen and Zhang Tian</i>	

Analysis of Science and Technology Activity Based on DEA Model	604
<i>Wenhai Gao and Lina Yuan</i>	

Information Management System

Modeling and Verifying the Kernel of RTOS	610
<i>ShengWen Gong</i>	

Improved Party Building Based on Information Technology	618
<i>Qiangqing Wang and Jierui Wang</i>	

Mobile Customer Satisfaction Evaluating System	624
<i>Yan Gao, Lihui Zhou, and Linfan Li</i>	

Performance Analysis of Transcritical CO ₂ Sewage Source Heat Pump by Visual Basic Program	630
<i>Hongli Wang, Huiqin Liu, and Jingrui Tian</i>	

Information Retrieval Services Based on Lucene Architecture	638
<i>Hang Li, Wanlong Li, Guochun Wang, and Xinyi Peng</i>	

Analysis on the Stability of Cavity Based on Cavity Monitoring System	646
<i>Zhang Haibo, Song Weidong, and Liu Fangfang</i>	

LED Package Based on MATLAB Program	653
<i>Zhiqin Huang, Xiuyun Zheng, and Fengling Liu</i>	

Design and Implementation of Fuzzy Query System Based on Ajax	659
<i>Fenghua Liu, Aiming Wang, Liang-Liang Chen, and Guang-Hui Wei</i>	

Basin Hydrometeorology Real-Time Forecast System	665
<i>Tao Peng, Cheng-zhi Wei, Jin-tao Ye, Jun-chao Wang, and Zhi-yuan Yin</i>	
Internet Health Resources Development Based on Geographic Information System	672
<i>Yun-Peng Zhang, Dan Li, and Ying Zhao</i>	
Research of Mobile Library Information Service	678
<i>Yun-hong Lv</i>	
Civil Explosives Distribution Management Information	686
<i>Xue-yu Mi, Peng Zhang, Lin Zhang, and Bo Dong</i>	
Research of the Pension Insurance Information Management System Based on the J2EE	694
<i>Ying Chen, Xiaosi Xu, Jun Zheng, and Hong Wang</i>	
Library Microblogging Based on Sina Microblogging Platform	702
<i>Yang Shuangqi, Xu Jing, and Shao Shengchun</i>	
Control Engineering and Applications	
XML-Based Web Data Pattern Discovery and Extraction	708
<i>Rui Jia, Shicheng Xu, and Chengbao Peng</i>	
Application of Decision-Making Model Based on Structure Entropy in IT-Outsourcing Supplier Selection	716
<i>Li Shi and Hong-bin Dong</i>	
Chinese Web Text Classification Model Based on Manifold Learning	722
<i>Shengli Shi, Zhibin Fu, and Jinzhao Li</i>	
Parallel Fault Information Mining Using Integrating Neural Classifier	729
<i>Hongsheng Su</i>	
Research of Knowledge Integration Based on Semantic Web for Drilling Risk Management	738
<i>Yingzhuo Xu</i>	
Transmission and Distribution Network Analysis Platform Based on FastDB	747
<i>Zhangzhuo Dong, Xiaoting He, Jianxin Li, and Kun Li</i>	
Subjective Testing System Based on Chinese Word Segmentation	756
<i>Jiao Cui-Ling and Wang Jian-Ping</i>	
An Efficient Algorithm for Reconstruction of Discrete Sets from Horizontal Projections with Absorption	762
<i>Chen Junyu, Gao Aili, and Zhang Chuanlin</i>	

Center-Based Iteration Algorithm of Pre-extracting Support Vectors	773
<i>Yun Shiwei and Shu Yunxing</i>	

Business Intelligence and Applications

Research on Inbound Tourist Market of Liaoning Province Based on Tourism Background Trend Line	783
<i>Wei Guo, Yixue Liu, and Hong Guo</i>	
Research on Influence of Tourism on Economy	789
<i>Dan Luo and Wuzhong Zhou</i>	
Hydrodynamic Characteristics Experiment of Vegetation Flexible Dam	795
<i>Feng Liu, Liang Ma, Liang-Liang Chen, and Guang-Hui Wei</i>	
Solar Energy Product Based on Variable Length Gene Expression and Intuition Models	802
<i>Ting-ting Zhao, Xiao-peng Wei, Zhi-qiu Xia, Hong-jiang Wang, and Jun-tie Xia</i>	
Study on the Logistics Efficiency of Three Northeast Provinces Based on Three-Stage DEA	810
<i>Sun Xu</i>	
Idols Supply Pattern in Confident Reestablishment of National Culture	820
<i>Shanguang Lv, Huifen Li, and Xiuxia Wang</i>	
Research of Popularizing Solar Heating in Countryside	827
<i>Junjun Wu, Bin Xia, and Mingchuan Meng</i>	
A Social Benefit Comprehensive Evaluation Model Based on Unascertained Measure Expectation of Science and Technology Information Products	833
<i>Jianli Chen</i>	
Utility Analysis on Influence of Monetary Policy Adjustment on Jiangsu Province Export Trade in International Financial Crisis	840
<i>Yue Ying, Ma Xing-hua, and Zhang Cheng-hu</i>	
Author Index	847

Table of Contents – Part II

Internet Computing and Applications

Research on User Authentication Scheme against the Man-in-the-Middle Attack	1
<i>Xiancun Zhou, Yan Xiong, and Renjin Liu</i>	
Study on Network Security Proactive Defense Model Based on Honey Pot Technology	9
<i>Haicheng Zhang</i>	
Framework of Server and Storage Virtualization	15
<i>Song Yibing and Cheng Jian</i>	
Efficient Management Scheme of Campus Information	21
<i>Jun Zheng, Fukai Cao, and Jiansheng Gao</i>	
Web Service Technology in College Financial Information Service System	27
<i>Jun Zheng, Jundong Tang, and Luyang Gao</i>	
Efficient Network Monitoring System	34
<i>Jiantao Gu, Yan Wu, and Zhao Gu</i>	
Software Design for Network Monitoring System	41
<i>Jiantao Gu, Benluo Xiao, and Zhao Gu</i>	
Network Content Safety Design of Monitoring System	47
<i>Jiantao Gu, Benluo Xiao, and Zhao Gu</i>	
Improvement Intrusion Detection Based on SVM	53
<i>Jian-hua Zhao and Wei-hua Li</i>	
Research on Subject Information Service Platform of University Library	61
<i>Lian Lijun</i>	
Networks Intrusion Behavior Prediction Based on Threat Model	67
<i>Dong Ma, Yongjun Wang, Feng Huang, and Zhenlong Fu</i>	
An Requirement-Aware Approach for Composing Web Services	76
<i>Xiangbing Zhou, Hongjiang Ma, and Fang Mao</i>	
Efficient Modeling Scheme Based on Smart Neural Network	85
<i>Yaping Yu, Jian Zhang, and Ying Li</i>	

Programming Design of Device Driver Based on Windows CE.Net.....	92
<i>Qiuhong Fan, Hao Zhang, Peng Li, and Jiongzhao Yang</i>	

Multimedia Networking and Computing

Contourlet-Based Fusion Method for Video Surveillance Using the Pulse Coupled Neural Networks Model	98
<i>Xi Cai, Guang Han, and Jinkuan Wang</i>	
Consensus-Based Detection Method for Visible and Thermal Videos	106
<i>Guang Han, Xi Cai, and Jinkuan Wang</i>	
M-Learning English Course Design in Chinese Local University	112
<i>Lin Hu</i>	
Multi-objective Optimization for Data Placement Strategy in Cloud Computing	119
<i>Lizheng Guo, Zongyao He, Shuguang Zhao, Na Zhang, Junhao Wang, and Changyun Jiang</i>	
Web-Based Teaching System of Tourism English	127
<i>Xiaohua Zhu</i>	
Explore Sharing Platform of Medical Educational Teaching Resources with IT Technology	133
<i>Deliang Wen and Xiaodan Ma</i>	
Groundwater Prediction Based on Time Series Model and Wavelet De-nosing	138
<i>Wei Yongxia, Qu Mingyang, and Wang Xiaoyan</i>	
Efficient Image Watermarking Algorithm Using Non-negative Matrix Factorization and Discrete Wavelet Transform	146
<i>Song Liang, Jinhui Lai, and Delong Cui</i>	
Study on Sleep Quality and Learning Outcomes of College Students	152
<i>Tian Jia-Li, Li Ji-Lin, and Niu Jian-Zhao</i>	
Study on Digital Library and User Interaction Platform	158
<i>Nan Pang</i>	
Extracurricular Reading of College Students Based on Statistics	164
<i>Jianying Miao</i>	
Intelligent Computing and Applications	
Code-Controlled Hardware Trojan Horse	171
<i>Hongbo Gao, Qingbao Li, Yu Zhu, and Yong Liu</i>	

Support Vector Machine Classification Algorithm and Its Application	179
<i>Yongli Zhang</i>	
A Flatness Pattern Recognition Model Based on Wavelet Transform and Probabilistic Neural Network	187
<i>Jiawei Wang, Ruicheng Zhang, and Wentao Song</i>	
Research on EEG Classification with Neural Networks Based on the Levenberg-Marquardt Algorithm	195
<i>Yue Chen and Shaobai Zhang</i>	
Research on Reduction Algorithm Based on Variable Precision Rough Set	203
<i>Wang Zongjiang</i>	
SOA Based MRO Planning System with Knowledge Repository	211
<i>Minglun Ren and Panpan Yang</i>	
A Modified Group Search Optimizer Algorithm for High Dimensional Function Optimization	219
<i>Lijin Wang, Xinxin Hu, Jing Ning, and Lin Jing</i>	
Research on Intelligence Optimization of Web Service Composition for QoS	227
<i>Jianjun Li, Bin Yu, and Wuping Chen</i>	
Decision-Making Scheme Based on LIOWAD	236
<i>Meihua Cheng and Shouzhen Zeng</i>	
Intuitionistic Fuzzy Decision Making Based on OWA and Distance Measures	243
<i>Shouzhen Zeng and Jiaqin Chen</i>	
Research on Fingerprint Identification in Mobile Telephone	249
<i>Fukai Cao, Zheng Zheng, and Yao Zhao</i>	
Autonomous Vision-Based Safe Area Selection Algorithm for UAV Emergency Forced Landing	254
<i>Aiyi Lu, Wenrui Ding, Jiaxing Wang, and Hongguang Li</i>	
Computational Statistics and Applications	
Non-extensive Statistical Mechanics and Statistical Distribution for Completely Open Systems	262
<i>Bin Yang, Heling Li, and Ying Xiong</i>	
The Convergence Theorems of Fixed Points for Nearly Uniformly <i>L</i> -Lipschitz Mappings	272
<i>Jian Zhang, Yaping Yu, and Yanli Liu</i>	

Explicit Iteration Scheme with Perturbed Mapping for Common Xed Points of a Nite Family of I -Asymptotically Nonexpansive Mappings	280
<i>Lingmin Zhang, Suhong Li, Xin Xiao, and Huijuan Zhao</i>	
On Distance Regular Graph of Valency 10 and $a_1=1$	288
<i>Junhong Ma and Nan Ji</i>	
Lie Symmetry Analysis for the Degasperis-Procesi Equation Based on Maple	293
<i>Qing Wang and Tao Wang</i>	
Oscillation Criteria for Nonlinear Neutral Perturbed Dynamic Equations on Time Scales	300
<i>Xiuping Yu, Hongkai Sun, Hongyu Yang, and Haibin Jing</i>	
An Analytical Method for Parameter Estimation Errors	309
<i>Qiuhong Fan, Qun Wei, Peng Li, and Baoxiang Liu</i>	
Inverse Signed Total Domination Numbers of Some Kinds of Graphs	315
<i>Zhongsheng Huang, Zhifang Feng, and Huaming Xing</i>	
Dynamic Construction of Additively Weighted Network Voronoi Diagram	322
<i>Xin Liu, Yili Tan, and Hongmei Yang</i>	
$(k,d)^*$ - Choosability of K_5 -Minor-Free Graphs	328
<i>Wenjing Miao</i>	
Approximation Theoretic Aspects of Probabilistic Representations for Bi-continuous Semigroups	334
<i>Dingbang Cang and Shigang Ge</i>	
Ruin Probability of Double Type Insurance Compound Negative Binomial Risk Model	341
<i>Wei Jing, Liu Haisheng, and Gui Wenyong</i>	
On a Class of Locally Dually Flat Isotropic Berwald Metrics	348
<i>Wenjing Zhao, Bochao Qu, Li Li, and Linan Shi</i>	
PDEs Solution Based on Two-Stage MFS-MPS and One-Stage MFS-MPS	356
<i>Xiaolei Wang, Tongsong Jiang, and Ziwu Jiang</i>	
On Explicit Determinants of RFPLR and RFPLL Circulant Matrices Involving Pell Numbers in Information Theory	364
<i>Zhao-Lin Jiang, Juan Li, and Nuo Shen</i>	
Solutions for One-Dimensional Time-Dependent Schrödinger Equations	371
<i>Tongsong Jiang, Xiaolei Wang, and Zhaozhong Zhang</i>	

Contact Relation and Way-Below Relation in Fuzzy Setting	379
<i>Xueyou Chen</i>	

Cloud and Evolutionary Computing

Integrated Model of Support Vector Machine Based on Optimization of Artificial Fish Algorithm	387
<i>Yong Ye, Lichuan Gu, and Shaowen Li</i>	
Improved Strategies of Ant Colony Optimization Algorithms	396
<i>Ping Guo, Zhujin Liu, and Lin Zhu</i>	
Improved Ant Colony Optimization for Detecting Functional Modules in Protein-Protein Interaction Networks.....	404
<i>Junzhong Ji, Zhipun Liu, Aidong Zhang, Lang Jiao, and Chunnian Liu</i>	
Trimmed Estimator of Multiple Linear Errors in Variables Model Based on Decimal Integer Encoding Genetic Algorithm	414
<i>Huirong Cao and Mengtian Yue</i>	
Cryptanalysis of Wang's Auditing Protocol for Data Storage Security in Cloud Computing	422
<i>Chunxiang Xu, Xiaohu He, and Daniel Abraha-Weldemariam</i>	
ADRC of Displacement Delayed Feedback System for the Rolling Mill Main Drive System.....	429
<i>Ruicheng Zhang, Pingping Yang, Peipei Wang, and Mingduo Li</i>	
Intrusion Detection Based on BP Neural Network and Genetic Algorithm.....	438
<i>Zhao Jian-hua and Li Wei-hua</i>	
Township Health Center Information System Based on Cloud Computing.....	445
<i>Huaiyong Nie and Lina Yue</i>	
Heuristic Algorithms of Single-Machine Multi-criteria Scheduling.....	451
<i>Shuming Guan, Junhong Ma, Xiaoqing Lu, and Wenyi Zhang</i>	
AIB-SPMDM: A Smartphone Malware Detection Model Based on Artificial Immunology	457
<i>Min Zhao, Tao Zhang, Jinshuang Wang, and Zhijian Yuan</i>	
Local Stable Mechanism for Particle Swarm Optimization Algorithm ...	466
<i>Bo Wei, Yuanxiang Li, Dingcai Shen, Fei Yu, and Xing Xu</i>	
Mode of Regional Library Alliance under Cloud Computing Environment	474
<i>Ye Suping and Yang Jing</i>	

Research of Grid Resource Scheduling Based on Improved Ant Colony Algorithm.....	480
<i>Dan Liu, Shi-xia Ma, Zu-hua Guo, and Xiu-lan Wang</i>	
Network Security Evaluation Model Based on Cloud Computing	488
<i>Jin Yang, Cilin Wang, Le Yu, Caiming Liu, and Lingxi Peng</i>	
A Quantum Potential Strategy by Quantum Genetic Algorithm	496
<i>Han Bing, Wang Shujun, and Ma Junhong</i>	
Circle-Based Improvement Strategy of Simulated Annealing Genetic Algorithm.....	502
<i>Han Bing, Jiang Junna, and Wang Xinchun</i>	
Experimental Platform Based on Cloud Computing	508
<i>Lina Yue, Huaiyong Nie, and Lu Dong</i>	
Alliance Mechanism Based on Cloud Computing	514
<i>Jing Xu, Shuangqi Yang, and Kui Yu</i>	
Computational Genomics and Proteomics	
High-Speed Multi-hop AMCP Algorithm in MWSN	521
<i>Dengyuan Xu and Hong Wang</i>	
Prediction of Rheumatoid Arthritis-Associated Antigen Peptides Using a Novel Mixed Algorithm.....	528
<i>Tao Liu, Jiuyun Zhang, and Tianyu Lei</i>	
Numerical Simulation Method for the SIV Epidemic Model with Impulsive Vaccination and Infection-Age	536
<i>Wei Wei</i>	
Dynamics and Simulations of Impulsive Semi-Ratio-Dependent Predator-Prey System with Functional Responses	544
<i>Wenxiang Zhang, Zhanji Gui, and Kaihua Wang</i>	
Orthogonal Design of Pharmaceutical Experiment Based on SPSS.....	552
<i>Ji Zhou, Ran An, Heng Zhang, and Yongqiong Liu</i>	
Optimization of Corn Furrow Irrigation Modes Based on RAGA-PPC Model	561
<i>Wei Yongxia, Zhao Yuyu, and Wang Xiaoyan</i>	
Research on Chinese Medicine Syndrome of Sleep Disorders of College Students	567
<i>Tian Jia-Li, Zhao Huiqin, and Li Cheng</i>	

Study on Treatment for Subhealth-Induced Insomnia among Student Patients by Self Point Massage	575
<i>Tian Jia-li, Zhang Ye-qing, and Han Jian</i>	
Evaluating System of Treatment of Quercetin and Resveratrol Mixture in Type 2 Diabetic Rats	583
<i>Yan Jiang, Zhilu Wang, Jinxiu Chu, Hong Yu, Ting Han, Xin Zhao, and Shuying Han</i>	
Medical Image Analysis in Treatment of Quercetin and Resveratrol Mixture on Type 2 Diabetic Cardiomyopathy Rats	590
<i>Yan Jiang, Jian Chang, Jing Bai, Hong Yu, Jinxiu Chu, Ting Han, Xin Zhao, and Shuying Han</i>	
Progressive Relaxation Training Assisted Cognitive Therapy of Insomnia	596
<i>Tian Jia-li, Zhao Hui-qin, and Xu Yan-mei</i>	
TCM Dyspnea Syndrome Experience Mining Based on Rough Set Theory	604
<i>Huixia Ma, Ya Wang, Wenjing Gong, and Jutai Bao</i>	

Knowledge Management and Applications

Fault Diagnosis of NPC Inverter Based on Multi-layer SVM	611
<i>Danjiang Chen, Yinzhen Ye, and Rong Hua</i>	
Study on Cashier Work of the Information Age	622
<i>Yinqiu Wang</i>	
Human-Computer Interaction in Ubiquitous Computing Environments	628
<i>Xianquan Zeng and Hongwen Pei</i>	
Research on Collaborative Decision Theory	635
<i>Aimin Wang, Xia Zheng, Ruihong Zhang, and Jie Li</i>	
Research on Knowledge Discovery Service System in the University Library	643
<i>Zhijiang Liao and Min Gao</i>	
A TNC Protocol for 3GPP-WLAN Integrated Environment	651
<i>Jun Ma and Yuanbo Guo</i>	
Risk Information Management of Complex Weapons System Development Based on Interpretative Structure Modeling	658
<i>Gao Shanqing, Wu Xiaoyun, Xing Lixin, and Wang Qiang</i>	

Demand Forecast of Regional Tourism Based on Variable Weight Combination Model	665
<i>Lijuan Liu</i>	
Optimal Algorithm for Maintaining Equipment Loading	671
<i>Su Yan, Pengcheng Yan, Jiangsheng Sun, Lianwu Zhang, and Shaoguang Wang</i>	
Document Modeling of PDM Based on STEP in Integrated Environment	678
<i>Wei Qun, Fan QiuHong, and Zhang Hao</i>	
Engineering Management and Applications	
Research on Copper Electroforming Based on Using of Nano-La ₂ O ₃	685
<i>Zhiyong Li, Zhipeng Duan, and Hongli Liu</i>	
A Memetic Algorithm Applied to Allocation Problem of the Concrete Mixing Plants	693
<i>Hao Zhi-feng, Wang Ai-Jing, and Huang Han</i>	
Research on Gascut-While-Drilling Volume Caculation Model during Underbalanced Drilling	701
<i>Jun He, Jing Li, Zhaoming Zhou, and Dong Zhang</i>	
Abhesion Mechanics Model of Horizontal Well Cement-Formation Interface	709
<i>Zhang Jinan, Wang Yinlong, Ai Chi, Yin Jindan, and Feng Fuping</i>	
Mechanical Behaviors of Shape Memory Alloys Concerning Small-Strain and Large-Strain Methods	717
<i>Jie Yao, Yong-hong Zhu, and Yun-zhang Wu</i>	
Orebody Model Compression Research Based on Decimal Morton Code	724
<i>Yajing Liu and Mei Li</i>	
Performance Analysis of Transcritical CO ₂ Compression Cycle	730
<i>Hongli Wang, Jingrui Tian, and Huiqin Liu</i>	
Research on Speed Estimation of PMSM Based on Radial Basis Function and MARS	737
<i>Wei Gao, Xiaojun Hua, and Zhirong Guo</i>	
Research on EEG Based on SVM and EMD	745
<i>Xinxin Wang and Jianlin Zhao</i>	
Digital and Analog Compatible TV Transmitter Power Measurement	752
<i>Sujuan Zhang, Yanli Zhang, Yongchang Liu, and Jun Ma</i>	

Evidence Theory-Based Safety Evaluation in Coalmine	759
<i>Liu Wensheng</i>	

Control Engineering and Applications

Study on Hopf Bifurcation of Displacement Delayed Feedback System for the Rolling Mill Main Drive System	766
<i>Ruicheng Zhang, Pingping Yang, Mingduo Li, and Xulei Hu</i>	
Quantitative and Spatial Layout Evolvement Model of Land Use Based on Fuzzy System	775
<i>Zuohua Miao, Yong Chen, and Xiangyang Zeng</i>	
Research on Web Text Mining	783
<i>Ruan Guangce</i>	
The Algorithmic Complexity of Reverse Signed Domination in Graphs	791
<i>Wensheng Li, Zhongsheng Huang, Zhifang Feng, Huaming Xing, and Yuejing Fang</i>	
SVM Classification Algorithm in ECG Classification.....	797
<i>Yanwei Zhu</i>	
Research on Flexible Management of Business Process	804
<i>Jiangjing Wang and Jiulei Jiang</i>	
The Research of Matrix Converter Space Vector Modulation.....	812
<i>Xu Zhuo, Zhao Shan, Chen Rui Min, and Xiao Yong</i>	
Optimization of Wheel Positioning Parameters of Automotive Front Suspension Based on ADAMS	821
<i>Lijun Qian and Qin Shi</i>	
Research on Equipment Failure Risk Control System Based on the Knowledge Integration	828
<i>Chen Yan-guang</i>	
Flow Control System Performance Optimization Base-On Zero-Copy ...	834
<i>Sujuan Zhang, Wei Li, Yanli Zhang, and Tao Wu</i>	
Author Index	841

WIPOMTS: An Internet Public Opinion Monitoring System

Shouhua Zhang¹, Kaiyu Shang², Shuai Cong³, Bin Zhang², and Zhenpeng Liu^{1,2}

¹ College of Mathematics & Computer Science, Hebei University, 071002 Baoding, China

² Network Center, Hebei University, 071002 Baoding, China

³ College of Industrial & Commercial, Hebei University, 071000 Baoding, China

{zhangshouhua, shang, congshuai, zb, lzp}@hbu.edu.cn

Abstract. Based on the characteristic of Internet public opinion occurrence and spreading, to improve the information acquisition strategy, and pursue key words to conduct topic clustering, to design the models for analyzing public opinion hotspots based on different features of news, forums and blogs. On this basis, an Internet public opinion monitoring system was designed and implemented. The running tests show that this system is capable of finding hot topics and tracking emergencies timely.

Keywords: Internet public opinion, topic clustering, hot topic, tracking.

1 Introduction

The data published by CNNIC in July, 2011 showed, up to the end of June, 2011, the Internet user has achieved 485million in china [1]. As a new form of information dissemination, Internet media has become part of people's daily life. The messages of netters have become more actively than ever before. Hot topics are constantly forming on Internet, many of which became social events, that reveals the assignable influence of Internet. It has caused wide attention that based on the acquisition of Internet public opinion, how to achieve public opinion collecting; hot topic finding, tracking and analyzing; information safety protection.

Study on Internet public opinion domestic and overseas has made some advantages, but there are still several problems to be resolved. The working range of current Internet public opinion monitoring and warning system is limited by the key words given by the user. Affected by some subjective factors such as knowledge, information source, and concerns of the user, the system will not detect those events that unexpected by the user [2]. Therefore, by using computers to finishing news, find hot topic key words automatically; update common safety word frequency in library timely, the timely tracking of sudden events can be completed [3]. At present, most hot tracking algorithms adopt Text Clustering technology. When dealing with mass web pages, with complicated contents and many disturbs, clustering causes huge central bias and needs to be improved [4].

The main advantage of this system is to improve the collecting strategy of public opinions, automatically mine key words of hot topics, put forward new analytical models, find hot topics timely, and improve efficiency.

2 System Architecture

Woodpecker Internet public opinion monitoring tracking system, abbreviated as WIPOMTS was developed. Based on the characteristic of Internet public opinion occurrence and spreading, the system improves the information acquisition strategy, and automatically mines key words from titles of web pages. It conducts topic clustering with these key words. By reading these topics, public opinion analysts can know what is exactly happening and what has happened. Furthermore, WIPOMTS can automatically and persistently track the event developments to assist the analysts rapidly, completely and comprehensively apprehending the general picture of an event. The system can find hot topics timely and track emergencies in real time.

WIPOMTS has five components, public opinion topic planning, public opinion information acquisition, public opinion information preprocess, public opinion information analysis and public opinion process.

(1) Public opinion topic planning refers that the public opinion supervising department, based on its needs, chooses proper public opinion topic and corresponding seed URL collections and then determines the acquisition task. A decided topic is the basis of public opinion analysis. It will be determined by applying the key words set in the system. Choice for sources of public opinion information is crucial to the next public opinion mining. Accurately grasping the original source of public opinion information, can obtain a more comprehensive public opinion.

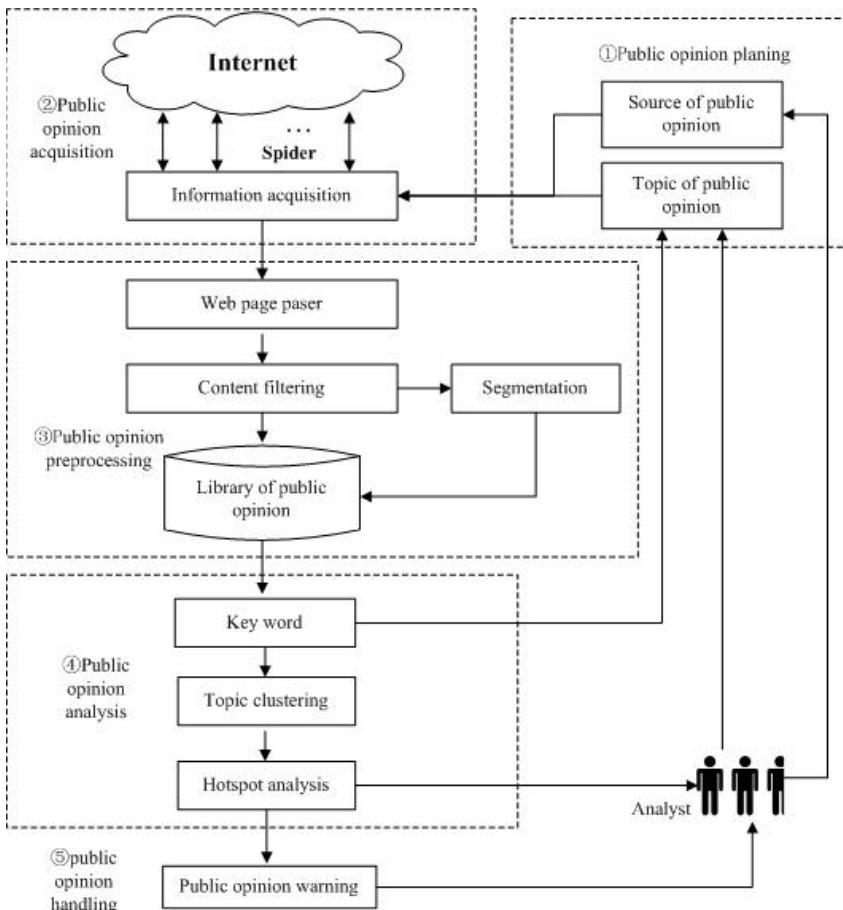
(2) Opinion information acquisition is a web acquisition process that the system automatically acquires information for Webs through linking relationships and extends to the whole Web by following the links. According to the different features of new, forum, blog, the system collects the different imformation points.

(3) Public opinion information preprocess is the preliminary processing or treatment on the acquired web information which lays the foundation for future processes. The preprocessing includes web page analysis, web text segmentation, words fliterring etc.

(4) Public opinion information analysis is the most important part in WIPOMTS, which includes hot topic mining on the acquired news, comments and other information according to the analytical model of hot topic; and warning analysis, text clustering, topic mergence etc.

(5) Public opinion process includes warning, reporting and guiding. First on the analysis result information warning and forecasting are performed according to the indicator, and then the regulated public opinion in reported to corresponding departments to provide auxiliary to decision-making.

System architecture is shown in Fig. 1.

**Fig. 1.** System Architecture

3 System Components

3.1 Public Opinion Topic Planning

A decided topic is the basis of public opinion analysis which will be determined by applying the key words set. Key words extraction is the clue of hot topics the quality of which diametrically determines the accuracy. Key words have the features of representativeness, conciseness, timeliness, mass information, high degree of association between key words, so that the topic and contents of the hot topic will maximally be covered by minimal information. There are two extraction methods, manual extraction and automatic extraction [5]. Automatic extraction refers to as the process of extracting common features from a set of web pages by the program, which will be weighted by frequencies. Manual extraction has the advantages of simple

implementation; human experiences are normally accurate so that huge deviation can be avoided. But there are several drawbacks such as absence and inaccuracy of weighting.

According to the hot topic distribution features, the system acquires information from timeliness mainstream media webs and searching engines, to guarantee that the information has reliable source, better timeliness, smaller quantity of information and shorter processing time [6].

3.2 Opinion Information Acquisition

Internet opinion information acquisition is the process of collecting of web pages related to the topic planning. The acquired web data is stored in the data base after preprocess to provide high quality data source to analysis department. A web queue is created and accessed via different protocols; finally the webs are downloaded for further analysis. The system employs multithreading parallel acquisition strategy to improve the efficiency.

To be concrete, all acquisition web sets are stored in the queue; then webs are ceaselessly assigned to threads to perform acquisition; when finished the thread, a request will be sent to the master process to assign a new web till the queue is empty.

3.3 Public Opinion Information Preprocess

Public opinion information preprocess includes web page analysis, segmentation, filtering etc [7]. The source web pages include many advertisements, pictures, links, which carries no value for the process and costs system resources and processing time. Meanwhile the data presents are based on different sources, and the system employs HTML analyzer syntax and regular expression to label the title, source, author, release time and text etc. Since the Chinese segmentation is well known, no more details are needed. After segmentation, the result will be filtered through the stop word list and filtering rules. The stop word list includes the function words such as auxiliary, preposition, conjunction etc. and no practical meaning single words [8]. For obvious no practical meaning strings, such as a great deal of numeral and classifier collocations, familiar and no meaning prefix and suffix, rules are designed to perform filtering.

3.4 Public Opinion Analysis

3.4.1 The Key Words Extraction

The web text related to Internet public opinion is different from ordinary web text. The title of this web text has a clear topic. In order to enhance the attractiveness of the web page and improve attention to the web page, the web page of Internet public opinion in general clearly expresses the topic of event and views in the title [9]. So, the key word can be extracted from the title of web text. According to the title we can get the vector representation of the document; $d = (t_1, t_2, \dots, t_i, \dots, t_m)$. t_i is a candidate key

word, which was handled with segmentation and filtering. m is the number of words. A candidate key word vector is built by the vector d of all the documents, the candidate key word vector as $t=(dt_1, dt_2, \dots, dt_n)$, dt_i is a candidate key word of some document. Then, the weight of the candidate key word is calculated in the vector t .

Each candidate key word includes TF, IDF, parts of speech and other information. Higher TF indicates more concerns; larger IDF means more obvious discrimination. Except TF and IDF, parts of speech and word length are also included [10]. In view of the parts of speech, a named entity uses more information than non-named-entity. Therefore the weight of named entity words is raised. Verbs are considered as the standard, other speeches are lower. Longer word presents more information. Therefore the calculation method of weight about key words is given in Equation 1.

$$Weight(dt_i) = tf(dt_i) \log_2(N / n_{dt_i}) * Weight(POS(dt_i)) * Length(dt_i)) \quad (1)$$

In the equation, $Weight(dt_i)$ is the weight of candidate key word dt_i , $tf(dt_i)$ is the frequency of dt_i . n_{dt_i} is the number of documents that contain dt_i in the title among the document sets. N is the total document's account. $Weight(POS(t))$ is the weight of t 's parts of speech, which is 2 for entity words, 1.5 for verbs and 1 for the rest. $Length(dt_i)$ is the length of dt_i . The key words are obtained by descending order the words.

3.4.2 Topic Clustering

Taking into account the influence of different sites and the timeliness of hotspot, the web pages based on the weight for the first factor, release time for the second factor, are sorted by the weight, time in descending order. First a keyword represents a hot topic is tolerated, and then it begin to cluster. The first key word in the key word set is regarded clue as the first hotspot topic. If the key word set of document includes this key word, the document is found out. All these documents found out by us are clustered. The first document is tolerated the first hotspot topic, and then we take out one topic from the other documents. We adopt cosine formulate to calculate the similarity degree between the topic and every hotspot topic. If all the values of similarity degree are less than the threshold P , we regard the topic as a new hotspot topic. Otherwise the topic is combined with the hotspot topic which is the largest similarity degree with the topic. We repeat the above steps for the remaining topics. Then we handle the remaining documents with the same steps according to the second key word in key word set. Algorithm executes iteratively, until all the documents are handled or the number of hotspot topic is enough.

3.4.3 The Analytical Model of Hotspot Topic

According to the different features of news, forum, Blog, we design the analytical models of hotspot topics respectively. At present, although there are many news sites, but they are uneven. The reliability of different websites and timeliness of news are different. News of the number of participants and the number of comments also reflect the heat value. Considering the above factors, the analytical model of hotspots about news is showed in Equation 2.

$$HotNews(t) = \sum_{i=1}^n Weight(S_i) * Weight(f(pn_i, cn_i)) \quad (2)$$

HotNews (t) indicates the news heat value of the topic *t*. *n* indicates the number of news on the topic *t*. *Weight(S_i)* indicates the weight of the website where is the news *i*. *Weight(f(pni,cni))* indicates the weight based on the number of participants and the number of comments. *pni* indicates the number of participants of the news *i*. *cni* indicates the number of comments of news *i*. $f(pni, cni) = \alpha * pni + \beta * cni, (0 < \alpha, \beta < 1, \alpha + \beta = 1)$, α and β are the adjustment coefficients, General $\alpha = 0.2$, General $\beta = 0.8$.

In addition to the useful information of news, there is other useful information for forums, the number of topic reproduced, for example. The analytical model of hotspots about forums is showed in Equation 3.

$$HotForumValue(t) = \sum_{i=1}^n Weight(S_i) * Weight(f(bn_i, rn_i)) * Weight(r_i) \quad (3)$$

HotForumValue(t) indicates the forum heat value of topic *t*. *Weight(S_i)* indicates the weight of the website where is the post *i*. *Weight(f(bn_i, rn_i))* indicates the weight based on the number of post *i* browsed and the number of post *i* replied. *bn_i* indicates the number of post *i* browsed; *rn_i* indicates the number of post *i* replied. $f(bn_i, rn_i) = \alpha * bn_i + \beta * rn_i, (0 < \alpha, \beta < 1, \alpha + \beta = 1)$, α and β are the adjustment coefficients, General $\alpha = 0.2$, General $\beta = 0.8$. *Weight(r_i)* indicates the weight of the number of post *i* reproduced.

The analytical model about blogs is similar to the analytical model about forums.

4 Conclusion

Above models were applied in WIPOMTS. The system was implemented by Microsoft visual studio .Net 2005. Considering the system deployment and ease of use, the system structure used B/S, and the DBMS used SQL Server 2005.

Based on the number of participants and comments, the weights in Equation 2 are given in Table 1. Based on the number of post browsed and the number of post replied, the weights in Equation 3 are given in Table 2.

Table 1. The weights based on the number of participants and comments

f(pn,cn)	Weights(f(pn,cn))
[0,500)	0.1
[500,1000)	0.2
[1000,2000)	0.3
[2000,3000)	0.4
[3000,5000)	0.5
[5000,7000)	0.6
[7000,10000)	0.7
[10000,15000)	0.8
[15000,20000)	0.9
[20000,∞)	1

Table 2. The weights based on the number of post browsed and post replied

f(bn,rn)	Weight(f(bn,rn))
[0,500)	0.1
[500,1000)	0.2
[1000,2000)	0.3
[2000,5000)	0.4
[5000,10000)	0.5
[10000,20000)	0.6
[20000,50000)	0.7
[50000,100000)	0.8
[100000,200000)	0.9
[200000,∞)	1

Fig. 2 shows the system interface.

Fig. 2. System Interface

Internet public opinion monitoring system has important applications background. A fast and effective Internet public opinion monitoring system was designed. The system automatically mines hot key words to conduct topic clustering. According to the different features of news, forum, Blog designs the analytical models of hotspot topics respectively. WIPOMTS can mine the hotspots of Internet public opinion automatically and timely. The system collects information from the authority sites

and search engines which have strong timeliness and high reliability, therefore the amount of information collection is small, so the speed of analyzing hot spots is increased. The sources of information have high reliability, so the hotspots are analyzed with high accuracy. Through the running of WIPOMTS, it shows that the system runs with high efficiency, and the Internet public opinion hotspots are recognized with high accuracy.

Acknowledgements. This work is partially supported by the National Nature Science Foundation of China (Grant No. 60873203), the National Philosophy and Social Science Foundation of China (Grant No. 10BTQ039) and the China Postdoctoral Science Foundation funded project (Grant No. 20070420700).

References

1. China Internet Network Information Center: 28th Statistical Report on the Internet Development in China. Technical report, CNNIC (July 2011)
2. Ding, J., Xu, J.: IPOMS: an Internet Public Opinion Monitoring System. In: International Conference on the Applications of Digital Information and Web Technologies, pp. 433–437. IEEE Press, London (2009)
3. Zheng, K., Shu, X., Yuan, H.: Hot Spot Information Auto-detection Method of Internet Public Opinion. Computer Engineering 36(3), 4–6 (2010)
4. Liu, H., Li, X.: Internet Public Opinion Hotspot Detection Research Based on K-means Algorithm. In: Tan, Y., Shi, Y., Tan, K.C. (eds.) ICSI 2010, Part II. LNCS, vol. 6146, pp. 594–602. Springer, Heidelberg (2010)
5. Pan, Z.: Research of Analyzing Public opinion on the Web Based on Topic Key Words. Journal of Suzhou University 25(5), 28–30 (2010)
6. Lu, B., Cheng, X., Chen, Z.: Overview of the Study of Internet Public Opinion Mining. Information and Documentation Services 2, 41–45 (2010)
7. Guan, Q., Ye, S., Yao, G., et al.: Research and Design of Internet Public Opinion Analysis System. In: IITA International Conference on Services Science, Management and Engineering (SSME 2009), pp. 173–177. IEEE Press, Zhangjiajie (2009)
8. Hua, B.: Stop-word Processing Technique in Knowledge Extraction. New Technology of Library and Information Service 8, 48–51 (2007)
9. Tang, H., Wang, H.: Application of Improved K-Means Algorithm to Analysis of Internet Public Opinion. Computer Systems & Applications 20(3), 165–168 (2011)
10. Lian, J., Liu, Y.: Web Data Preprocessing and Automatic Abstract for Internet Public Opinion. Journal of Beijing Jiaotong University 34(5), 94–99 (2010)

Social Context Enabled Description Model for Web Services

Xiao Zheng, Qingyong Wu, Dongliang Ke, Haixia Li, and Yaqing Shi

School of Computer Science, Anhui University of Technology, 243002 Maanshan, China
zhx@ahut.edu.cn

Abstract. In pervasive environments, efficient automatic service discovery and composition depends on the information of service context, however, existing service description models lack of the description of social relationships. This paper analyzed the parameters of the social context of services from three aspects, i.e. the service requester, service provider and relationships between them, and proposed a semantic Web service description model which supports the description of social context. In order to extend the OWL-S metamodel to the S-OWL-S(SCA-enabled OWL-S) metamodel, Service Social Context Attribute (SCA) was added into the OWL-S. Finally, a case study is used to validate the S-OWL-S in service discovery.

Keywords: Pervasive Computing, Social Context, Semantic Service Description, S-OWL-S model.

1 Introduction

With the smart phone, PDA, and other embedded mobile devices become more and more popular in internet, how to effectively share a variety of services with service requesters is an important issue[1]. In a service's interactive process, it not only involves its own functionality, quality of service (QoS) but also is related to some relationship between the service provider and service requester, which is generated from spatial, temporal, requester hobby and users' evaluation, etc. Therefore, these various links in the interaction constitute a kind of social relationship. In order to optimize service discovery, service composition and its interoperability in pervasive computing, it is necessary to study the social relationship of services[2,3]. In [4], the concept of service context is proposed, which is any information that can be used to characterize the situation of an entity. An entity is any person, place or object which considered in the interaction between a user and a service as well as among services, including the users and services themselves.

Current research has considered context as a part of the service feature and applied it into service discovery and composition. In order to use the service context describing service feature, it is necessary to build a service context model and design a service description language supporting service context. C. Lee *et al* characterized the dynamic feature of service using context [5]. C. Dorn *et al* [4] defined the context dominance (CD) concept that allows some context attributes with higher priority, which will be considered prior to others. Standard service description language

(e.g., WSDL[6], OWL-S[7]) only describe service functionality, interface and executive process. In order to enhance the description capability of the service model, some work has extended standard language, for example, literature [8] proposed an effective extension of the WSDL that adds QoS feature. Although they have done a better improvement of service description, but still can not meet the requirement of service requester to accurate and fast lookup of the target service in pervasive computing environments. In order to enable user to quickly get the target service, it is necessary to consider the specific environments in pervasive computing.

Service discovery and composition need to consider social relationships under the pervasive environments. This paper introduces a semantic Web service description model to support the description of social context, which extends the semantic Web service description language OWL-S to support the description of social context. This paper proposes Social Context Attribute (SCA) as a part of the OWL-S Profile class. By OWL-S extension, SCA supports service lifecycle management in pervasive environments, and takes advantage of semantic Web service in service discovery, selection, and composition.

The remainder of the paper is organized as follows. Section 2 introduces social context of services and illustrates the proposed S-OWL-S metamodel. Section 3 gives a case study of S-OWL-S application. Finally, section 4 concludes our research contributions and future work.

2 Social Context of Service and Extended OWL-S Model

2.1 Social Context of Service

In the service interaction process, we must be able to describe the provider's services and the services requester needed. Literature [9] described the context ontology of user and service which solve the problem of service selection in pervasive computing environments. Literature [10] built the common social context ontology in pervasive environments. In fact, different domain requires different social context. In this paper, the social context of service is considered from the following three aspects.

Firstly, consider the requester's own social attributes.

Preference: The service requester's preference feature is often not easy to change, and influences the requester to choose services. For example, according to the requester ordinarily prefers comedy, horror, or action films to choose the movie service which you want to watch.

mood: The requester's mood is an important factor affecting the required service. This paper proposes the mood metrics: joy, anger, sadness, boredom, and others, which is an enum type, It can be determined by the user's choice. Of course, user may not sure how to express his feeling, but you can select a value to find some services. For example, if a user feels joy in a time, you can recommend him to do online ordering service or do online shopping service.

identity: The requester's identity or social status is an important reference factor of access control. Some services allow only some specific people or programs to access, such as different levels of registered members use different services. Simultaneously, it also represents the requester's preference with different types of services, such as teachers may be more interested in education service.

Secondly, consider the service program's social attributes.

domain: According to different industry domain, services can be divided into their respective categories, i.e. tourism, catering, etc. Reasonable classification helps to improve the efficiency of service discovery.

time: Provided services are time-related. Such as data backup services often run in the weekend.

location: Provided services are effective within some scope. Such as parking service is directly related to a location.

Thirdly, consider the social relations in interaction between service providers and requesters.

trust: A degree of trust or dependence generated in the interaction of service requester and provider. Requester depends on the level of confidence to choose provider's services. The feature can quantify trust with an integer from 0 to 5, 0 indicates mistrust, and 1 to 5 indicates the level of trust.

evaluation: Consumers give a comprehensive assessment on the service which they used. It's on behalf of satisfaction degree of service quality. It is calculated as follows:

$$E = \begin{cases} 0 & T = 0 \\ \frac{\sum_{i=1}^n e_i}{n} & T = 1 \end{cases} \quad (1)$$

where T is the truth of service description, $e_i \in [0,1]$ denotes a user evaluation value, and n is the number of user evaluation. The truth of service description is a primary factor which affects the evaluation of Web service. When $T=0$ that is the service provider is a fraud, the evaluation value of the service is 0, completely dissatisfied. When $T=1$ that is the service description is real, they use the average user evaluation to measure the evaluation value of the service.

According to the above analysis, social context consists of multiple properties. Each property has its own value and type.

Definition 1: Social context attribute $SCA = \langle id, value, type \rangle$, where id denotes the identifier of SCA ; $value$ denotes the value of SCA ; $type$ denotes the type of the value of SCA .

Definition 2: Web Services Description Model $WS = \{S, F, QoS, SCA\}$, where S is a basic description, namely the service name and text description; F is a functional description of the service, namely functional description and behavior description; QoS is quality of service, including cost, response time, reliability and reputation degree; SCA is the proposed social context.

2.2 Ontology Structure of S-OWL-S

OWL-S supports semantic service description, but does not support service context. Therefore, this paper proposes extending the OWL-S ontology to the S-OWL-S ontology in order to achieve the support of context description.

In OWL-S, the basic information of service description has three ontology, i.e. ServiceProfile, ServiceModel and ServiceGrounding. ServiceProfile describes what the service does; ServiceModel describes a client how to use the service; ServiceGrounding describes how an agent can access a service. Context is the dynamic information of service, so it is included in ServiceProfile. The class of SocialContextAttribute is added into ServiceProfile, and constitutes S-OWL-S ontology.

2.3 S-OWL-S Metamodel

Social Context Attribute (SCA) is designed as a class, called SCA class, and considered as a part of the OWL-S Profile ontology. Social context attributes are described by SCA class's properties. Profile ontology uses XML Schema structure, and the portion of the SCA class is as follows.

```

<owl:ObjectProperty rdf:ID="hasSCA">
  <rdfs:domain rdf:resource="#Profile"/>
  <rdfs:range rdf:resource="#SocialContextAttribute"/>
</owl:ObjectProperty>
<owl:Class rdf:ID="SocialContextAttribute"/>
<owl:DatatypeProperty rdf:ID="evaluation">
  <rdfs:domain rdf:resource="#SocialContextAttribute"/>
</owl:DatatypeProperty>
<owl:Class rdf:about="#SocialContextAttribute">
  <rdfs:comment>.....</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#evaluation"/>
      <owl:cardinality
rdf:datatype="&xsd;#boolean">1</owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<owl:DatatypeProperty rdf:ID="time">
  <rdfs:domain rdf:resource="#SocialContextAttribute"/>
</owl:DatatypeProperty>
<owl:Class rdf:about="#SocialContextAttribute">
  <rdfs:comment>.....</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#time"/>
      <owl:cardinality
rdf:datatype="&xsd;#date">1</owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

In order to support the visual edit of S-OWL-S, we develop a visual editor tool which is based on the ontology editor and knowledge acquisition tool Protege_4.1 and the OWL ontology visualization tool OWLGrEd. The UML format of extended profile ontology.

Extended Profile is divided into the following four parts:

Describes the property of connection Service class. Service class is a superclass, and the reciprocal properties of presents and presentedBy describe the relation of the Service and Profile.

Describes the basic information of the service, such as service name, service text description.

Describes the functionality characteristics of the service. Profile ontology uses hasInput, hasOutput, hasPrecondition and hasResult to describe the service's Input, Output, Precondition, Result (IOPR). In fact, the instance of IOPR are created in the Process, Profile instance is the mapping of Process instance.

Describes the non-functional characteristics of the service. serviceParameter, serviceCategory respectively describe quality of service, service category, as well as extension attribute (hasSCA) describe the social context.

3 Case Study

This section illustrates the important role of the suggested S-OWL-S in process of service discovery by the example of teaching materials services in a digital campus scenario. The teaching materials services S_t and S_w are respectively deployed in the teacher's smart phone and the campus website, which have the same functionalities. The key of service discovery technique is to recommend more appropriate services to service requesters. In this paper, we extend Protege's OWL-S Editor plug-in to support Web service description based on S-OWL-S.

(1) generate the Profile documents of S_t and S_w . Due to the limited space, there is only S_t 's social context listed below.

```
<profile:SocialContextAttribute
rdf:id="SocialContextAttribute">
<profile:domain rdf:datatype="&xsd;#enum"
>Education</profile:domain>
<profile:time>
<time:begin rdf:datatype="&xsd;#time"
>8:00AM</time:begin>
<time:end rdf:datatype="&xsd;#time"
>17:00PM</time:end>
</profile:time>
<profile:location>
<location:longitude rdf:datatype="&xsd;#int"
>119</location:longitude>
<location:latitude rdf:datatype="&xsd;#int"
>40</location:latitude>
```

```

</profile:location>
<profile:trust rdf:datatype="&xsd;#enum"
>5</profile:location>
</profile:SocialContextAttribute>
```

The XML-based code describes the profile of the teaching materials service generated by our extended OWL-S Editor, which is as a part of S_t Profile document. S_t 's domain is education, its service time is class time (e.g., 8:00 AM ~ 17:00 PM), its location can be obtained by GPS (e.g., longitude 119 degrees, latitude 40 degrees), and its trust is 5 that means its services can be absolutely trusted. Simultaneously, deployed on the campus website, S_w 's domain is education, its service time and location is unlimited, and its trust is also 5.

(2) generate the service requester's requirement description. Service requester can use smart phones, PC and other equipments to access teaching materials service. Suppose that there is a student Stu who access the teaching materials service through his smart phone, and his social context description is as follows:

```

<profile:SocialContextAttribute
rdf:ID="SocialContextAttribute">
<profile:location>
<location:longitude rdf:datatype="&xsd;#int"
>120</location:longitude>
<location:latitude rdf:datatype="&xsd;#int"
>50</location:latitude>
</profile:location>
<profile:identity>
<identity:type rdf:datatype="&xsd;#enum"
>Student</identity:type>
<identity:ID rdf:datatype="&xsd;#int"
>20090248</identity:ID>
</profile:SocialContextAttribute>
```

The identity denotes his major and grade, and the location denotes his movement area, which can be gotten by the smart phone's GPS.

(3) social context-aware service discovery. Although S_t and S_w have same functionalities, different access cost in different environments, e.g., S_t can be free access via WiFi, and S_t must be access via the payment gateway. According to the social context, the student Stu wants to select a more economical way to access, which is the key issue in service discovery. When Stu and his teacher are both in a same classroom, and the teacher can provide service at the moment, Stu can directly access the service S_t via WiFi. Because the cost of access the service S_t is less than S_w which is on the campus website, S_t is recommended to Stu .

This example selects the most appropriate service via location context, and other social context will not discussed any more. In a word, under the conditions of the same service functions, service discovery results can be optimized by social context.

4 Conclusions

Effective service description model is the basis of automatic service discovery and composition. This paper analyzes the social context emerged from services interaction under the pervasive computing environments, and proposes the Semantic Web service description model which supports the social context. We extend OWL-S to support the description of the social context. The model can be applied to a service discovery engine to enhance the efficiency of service discovery. This paper provides the property of the social context, but it may be not enough. The principle of determining context should be conducive to optimize the mechanisms of service discovery or composition. Our next step is to further study service interactions, and propose social context-aware service discovery and composition algorithms.

Acknowledgments. This work is supported by National Natural Science Foundation of China under Grants No. 61003311, Jiangsu Provincial Key Laboratory of Network and Information Security Grants No. BM2003201-201006, Anhui Provincial Higher Education Excellent Young Talents Foundation of China under Grants No. 2010SQRL040, Anhui Provincial Natural Science Research Key Project of China under Grants No. KJ2011A040. Anhui University of Technology 2010 Graduate Innovation Research Foundation of China under Grants No. 2010026.

References

1. Luo, J.Z., Wu, W.J., Yang, M.: Mobile Internet: Terminal Devices, Networks and Services. Chinese Journal of Computers 34, 2029–2051 (2011)
2. Faci, N., Maamar, Z., Abdeldjelil, H., Benslimane, D.: Towards a Framework for Weaving Social Networks Principles into Web Services Discovery. In: 11th Annual International Conf. on New Technologies of Distributed Systems, pp. 1–8. IEEE Press, New York (2011)
3. Maaradji, A., Hacid, H., Daigremont, J., Crespi, N.: Towards a Social Network Based Approach for Services Composition. In: 2010 IEEE International Conference on Communications, pp. 1–5. IEEE Press, New York (2010)
4. Dorn, C., Dustdar, S.: Sharing Hierarchical Context for Mobile Web Services. Distributed Parallel Database 21, 85–111 (2007)
5. Lee, C., Helal, S.: Context Attributes: An Approach to Enable Context-awareness for Service Discovery. In: 2003 Symposium on Applications and the Internet, pp. 22–30. IEEE Press, New York (2003)
6. Web Services Description Language Version 2.0,
<http://www.w3.org/2002/ws/desc/wsdl20-primer/>
7. Semantic Markup for Web Services, <http://www.w3.org/Submission/OWL-S/>
8. Ambrogio, A.D.: A Model-driven WSDL Extension for Describing the QoS of Web Services. In: IEEE International Conf. on Web Services, pp. 789–796. IEEE Press, New York (2006)
9. Zhang, L.C., Zou, H., Yang, F.C.: Multi-Attribute Group Decision Making-Based Decentralized Web Service Selection. In: 2010 International Conf. on Service Sciences, pp. 82–87. IEEE Press, New York (2010)
10. Biamino, G.: Modeling Social Contexts for Pervasive Computing Environments. In: 2011 IEEE International Conf. on Pervasive Computing and Communications Workshops, pp. 415–420. IEEE Press, New York (2011)

Improved Learning Algorithm for Self-adaptive Neural Nets Based on Principal Component Analysis

Shujie Du

Computing Centre, Ocean University of China 266100 Qingdao, Shandong, China
qingdaobeach@gmail.com

Abstract. This paper presents an analysis of the reason considering principal component analysis as a quantitatively rigorous method to remove redundant information from source cube, and gives the mathematical equation derivation for solving single and multiple principal components in multiple input and multiple output nets, puts forward the learning algorithm deduction of self-adaptive principal component feature extraction in neural network. The validity of the presented algorithm is demonstrated with simulated data at last.

Keywords: PCA, dimensionality reduction, self-adaptive feature extraction, eigenvector, convergence.

1 Introduction

In data sets with many variables, groups of variables often move together. One reason for this is that more than one variable may be measuring the same driving principle governing the behavior of the system. In many systems there are only a few such driving forces. But an abundance of instrumentation enables us to measure dozens of system variables. When this happens, we can take advantage of the information redundancy, simplify the problem by replacing a group of variables with a single new variable. Principal components analysis (PCA) is a quantitatively rigorous method for achieving this simplification. The method generates a new set of variables called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of data.

PCA aims at finding a set of M orthogonal vectors, which show variances diametrically as far as possible, and project them from original N-dimensional space to this M-dimensional orthorhombic subspace ($M < N$), so dimension is cut down consequently. To be highlighted, the first PC (abbr. principal component) orients to the direction of the largest variance, the second PC is limited in the subspace perpendicular to the first PC, it also moves along the direction of maximum variance in the subspace. Later PCs put in order according to the rule, PC_k follows to the direction of eigenvector corresponding to the k largest eigenvalue of covariance matrix. To put it another way, PCA corresponds to some optimal linear transformation $y = Px$, the optimal

component x and transformation P are based on making mean-square errors of y minimum. Supposed $R = E[xx^T]$ as correlation matrix, Row vectors of P correspond to eigenvectors of R 'maxeigen, normalized row vectors of P is PC with random input process [1].

2 Algorithm Solving Single PCA

As for single PCA extractor, Artificial neural network(ANN) is usually used to solving problems for given patterns of random data sets. The following design idea raised by simulating Oja neuron activities. Assumed that the neuron input and output are x_1, x_2, \dots, x_n and y respectively.

$$y = W^T x \quad (1)$$

y could be regarded as the projection on maximum variance direction. Object function is defined as:

$$f(W) = \frac{E[y^2]}{W^T W} = \frac{W^T R W}{W^T W} \quad (2)$$

W is primary eigenvector of Matrix R once $f(W)$ reaches a maximum, and $f(W)$ is thus the maxeigen. Gradient of function f is:

$$\nabla_W f = \frac{2RW(W^T W) - (W^T RW)2W}{(W^T W)^2} \quad (3)$$

Allowing for constraint condition $W^T W = 1$, Then comes to the conclusion,

$$\nabla_W f = 2E[xx^T]W - 2E[y^2]W \quad (4)$$

Without regard to expected factor, by using instantaneous value, random algorithm for updating weight vector W could be figured out as follows,

$$\Delta W = \eta \nabla_W f = \eta(yx - y^2 W) = \eta y(x - yW) \quad (5)$$

In fact, the problem above can also be described as following optimization problem.

$$\max\{f(W) | W^T W = 1\} \quad (6)$$

Note that if $R^T = R$, its eigenvalues are real and eigenvectors are Mutual Orthogonal; if R is positive semidefinite, all eigenvalues are positive or zero [2].

Lemma 1. According to the updating algorithm described in Eq. (5), once neural network comes converged, we can get the conclusion:

$$(1) \|W\| = 1 \quad (\sum_{i=1}^n W_i^2 = 1)$$

(2) W is along the maximum eigenvector of R , which also make $E[y^2]$ maximized.

Proof. Based on Eq.(5), once neural network comes converged, mean value of $\Delta W = 0$, characteristic equation is as follows,

$$0 = \frac{E[\Delta W]}{\eta} = E[yx - y^2 W] = E[xx^T W - W^T xx^T WW] = RW - [W^T RW]W \quad (7)$$

Make $RW = \lambda W$, $\lambda = W^T RW = W^T \lambda W = \lambda \|W\|^2$, So we could draw the conclusion:

$$\|W\|^2 = 1$$

Secondly, we will prove W is along the maximum eigenvector of R . Make x^α is a normalizing eigenvector,

$$Rx^\alpha = \lambda_a x^\alpha, \quad \|x^\alpha\| = 1$$

Supposed once network comes converged, W approximates x^α , $W = x^\alpha + \varepsilon$, ε is disturbed value, according to Eq. (7),

$$\begin{aligned} E[\Delta \varepsilon] &= E[\Delta W] = R(x^\alpha + \varepsilon) - [(x^\alpha + \varepsilon)^T R(x^\alpha + \varepsilon)](x^\alpha + \varepsilon) \\ &= R\varepsilon - 2\lambda_a [\varepsilon^T x^\alpha] x^\alpha - \lambda_a \varepsilon + O(\varepsilon^2) \end{aligned} \quad (8)$$

Now compute the projection of $E[\Delta \varepsilon]$ on axis of another normalizing eigenvector x^β , premultiply Eq. (8) with x^β , and omit $O(\varepsilon^2)$,

$$\begin{aligned} (x^\beta)^T E[\Delta \varepsilon] &= \lambda_\beta (x^\beta)^T \varepsilon - 2\lambda_\beta [\varepsilon^T x^\alpha] \delta_{\alpha\beta} - \lambda_\alpha (x^\beta)^T \varepsilon \\ &= [\lambda_\beta - \lambda_\alpha - 2\lambda_\alpha \delta_{\alpha\beta}] (x^\beta)^T \varepsilon \end{aligned} \quad (9)$$

If $\alpha \neq \beta$, equation(9) is simplified to the follows,

$$(x^\beta)^T E[\Delta \varepsilon] = [\lambda_\beta - \lambda_\alpha] (x^\beta)^T \varepsilon \quad (10)$$

In accordance with inner product $\langle \varepsilon, x^\beta \rangle$, Eq.(10) evolves in two conditions.

If $\langle \varepsilon, x^\beta \rangle > 0$, which means the projection of ε on axis of x^β is positive, the right side of Eq. (10) will be negative while $\lambda_\beta < \lambda_\alpha$.

If $\langle \varepsilon, x^\beta \rangle < 0$, which means the projection of ε on axis of x^β is negative, the right side of Eq. (10) will be positive while $\lambda_\beta < \lambda_\alpha$.

In summary, $E[\Delta \varepsilon]$ changes toward the positive direction of x^β , that is to say, W tends to change toward some larger eigenvector x^α , So once the network comes converged, W would be located along the direction of maximum eigenvector of R . and derived from $E[y^2] = W^T RW$, quadratic $W^T RW$ will reach a maximum.

The proved lemma shows that according to the rule given by Eq. (5), Weight vector \mathbf{W} will converge to the first PC, which corresponds to the normalizing eigenvector of \mathbf{R} 's maxeigen [3].

3 Improvement of Solution for Multiple PCA in MIMO Nets

By the above method, We can only figure out the first PC, if multiple input and multiple output (MIMO) network is concerned, weight vectors will converge to the same projection, which means all output vectors have the same result. So new algorithm should ensure variance maximized and the weight vectors Orthogonal, outputs will thus uncorrelated, eigenvector can be figured out by the way of solving restricted problem with Lagrange multiplier [4]. Take two outputs for example, the second output vector could be expressed as $y_2 = W_2^T x$, the corresponding objective function can be expressed as Eq. (11), which is aimed at searching optimum \mathbf{W}_2 .

$$f(W_1, W_2) = \frac{W_2^T R W_2}{W_2^T W_2} + \lambda(W_1^T W_2) \quad (11)$$

\mathbf{W}_2 should meet the condition of constraint, $W_2^T W_1 = 0$. The process of searching for the optimal solution equals to the description: $\max\{f(W_1, W_2) | W_2^T W_1 = 0\}$. Gradient referred to \mathbf{W}_2 is shown as:

$$\nabla_{w_2} f = \frac{2RW_2(W_2^T W_2) - (W_2^T RW_2)2W_2}{(W_2^T W_2)^2} + 2\lambda W_1 \quad (12)$$

Make $\nabla_{w_2} f = 0$, so we get the equation $\lambda = -E[y_1, y_2]$, also we know that $W_2^T W_2 = 1$, with these expressions Eq. (12) can be simplified as,

$$\nabla_{w_2} f = 2E[xx^T]W_2 - 2E[y^2]W_2 - 2E[y_1y_2]W_1$$

Take the instantaneous to mathematical expectation, comes the conclusion:

$$\Delta W_2 = \eta \nabla_{w_2} f = \eta(y_2x - y_2^2W_2 - y_1y_2W_1) \quad (13)$$

Similarly, as to the condition of m outputs, the updating rule of \mathbf{W}_i can be expressed as:

$$\Delta W_i = \eta(y_i x - y_i^2 W_i - \sum_{j=1}^{i-1} y_j y_j W_j) \quad \forall i = 1, 2, \dots, m \quad (14)$$

Different from Eq. (5), in order to ensure weight vectors are orthogonal each other, $\sum_{j=1}^{i-1} y_j W_j$ should be taken away from input vector x when calculating \mathbf{W}_i .

According to the algorithm of solving single PCA problems, and similar to the steps for proving lemma1, we get the following conclusion.

Lemma 2. Based on the algorithm shown in Eq.(14), Once network becomes converged, m output weight vectors would be on axes of the first m PCs of R. In other word, $W_i \rightarrow \pm x^i \quad i = 1, 2, \dots, m$, x^i corresponds to the normalizing eigenvector of eigenvalue λ_i in correlation matrix R, and the eigenvalues collate in descending order.

4 Learning Algorithm of Self-adaptive Principal Component Feature Extraction

Making reference to a self-adaptive PC feature extraction neuron model, whose input is $\{x_1, x_2, \dots, x_n\}$, output is $\{y_1, y_2, \dots, y_m\}$, weight matrix is $P=[P_{ij}]$. Supposed that the random input procedure is stable and there exist n different positive eigenvalues in autocorrelation matrix $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. In the mean while, supposed that the former m-1 output neurons have been converged to the former normalizing PC of input sequence. The most important feature of self-adaptive PC feature extraction lies in that neuron m chooses the max vector component orthogonal to the former m-1 components, which have been expressed by m-1 trained neurons [5].

As we know, activation transfer function of each neuron is linear combination of its input [6].

$$\begin{cases} y = Px \\ y_m = px + Wy \end{cases} \quad x = [x_1, \dots, x_n]^T, y = [y_1, \dots, y_{m-1}]^T \quad (15)$$

P is the weight matrix of former m-1 neurons, p is the row vector of weight p_{mj} . Learning algorithm of neuron m is::

$$\begin{cases} \Delta p = \beta(y_m x^T - y_m^2 p) \\ \Delta W = -\gamma(y_m y^T + y_m^2 W) \end{cases} \quad (16)$$

β, γ are parameters with different learning rates. Specific to each weight p_{mj} ,

$$\begin{cases} \Delta p_{mj} = \beta(y_m x_j - y_m^2 p_{mj}) & j = 1, 2, \dots, n \\ \Delta W_j = -\gamma(y_m y_j + y_m^2 W_j) & j = 1, 2, \dots, m-1 \end{cases} \quad (17)$$

Because the former m-1 neurons have been converged, connection weights between each other tend to 0. So W_m will converge to PC_m . In fact, $-\gamma y_m^2 W_j$ in Eq. (17) takes a negative effect on y_{m-1} by adding item $y_m W_j$ [7,8].

Supposed neuron 1 to m-1 have converged to the former m-1 PCs, $P = [e^1 e^2 \dots e^{m-1}]^T$, e^1, \dots, e^{m-1} are the former m-1 normalizing eigenvectors of R,

make $p(t) = \sum_{i=1}^m \theta_i(t)(e^i)^T$, t represents the training round, m represents the input patterns [9].

Lemma 3. Based on the assumption above, guided by the learning algorithm given in Eq. (17), once network becomes converged, W_m will come converged to the PC_m of R.

Proof. The proving can be divided into two parts. Firstly, the analysis on the former m-1 PC. Secondly, analysis on the new m,...,n PC.

According to Eq. (16), supposed $p(t)$ tends approximately to be a constant during one training round, we can get

$$p(t+1) = p(t) + \beta'[(p(t) + W(t)P)R - \sigma(t)p(t)] \quad \begin{cases} \sigma(t) = E[y_m^2(t)] \\ \beta' = m\beta \end{cases} \quad (18)$$

For simplicity, dedicating to characteristic patterns, updating rules to θ_i can be written as follows,

$$\theta_i(t+1) = [1 + \beta'(\lambda_i - \sigma)]\theta_i(t) + \beta'\lambda_i W_i(t) \quad (19)$$

Similarly, Eq. (16) changes to

$$W_i(t+1) = -\gamma' \lambda_i \theta_i(t) + [1 - \gamma'(\lambda_i + \sigma)]W_i(t) \quad \gamma' = m\gamma \quad (20)$$

Put two equations together, written in vector matrix as follows,

$$\begin{bmatrix} \theta_i(t+1) \\ W_i(t+1) \end{bmatrix} = \begin{pmatrix} 1 + \beta'(\lambda_i - \sigma) & \beta' \lambda_i \\ -\gamma' \lambda_i & 1 - \gamma'(\lambda_i + \sigma) \end{pmatrix} \begin{bmatrix} \theta_i(t) \\ W_i(t) \end{bmatrix}$$

When $\beta = \gamma$, matrix has a double eigenvalue, $p_i(t) = 1 - \beta' \sigma(t) < 1$. β' is a tiny positive. All θ_i, W_i tend to 0 at same rates because eigenvalues of all patterns are the same. Properly selecting β and γ is very important to reach some appropriate convergence rate. In fact, we usually make $\beta = \gamma = \frac{1}{m\sigma}$, and σ is adopted as the average value of y_m^2 .

As for proving on the second part, we can only take pattern i ($i \geq m$) into account, for W is just concerned with the former m-1 PCs, so with regard to $i \geq m$,

$$\theta_i(t+1) = [1 + \beta'(\lambda_i - \sigma)]\theta_i(t) \quad (21)$$

When $i = 1, 2, \dots, m-1$, θ_i and W_i will come converged to 0, thus we have,

$$\sigma(t) = \sum_{i=m}^n \lambda_i \theta_i^2(t) \quad (22)$$

Whatever θ_i changes, there exists $1 + \beta'(\lambda_i - \sigma) < 1$. For θ_i reduces gradually, Eq. (21) would not become diverged. Supposed that $\theta_m(0) = 0$, and define

$$\alpha_i(t) = \frac{\theta_i(t)}{\theta_m(t)} \quad i = m+1, \dots, n \quad (23)$$

Make Eq. (21) substituted into Eq. (23),

$$\alpha_i(t+1) = \frac{1 + \beta'(\lambda_m - \sigma)]}{1 + \beta'(\lambda_i - \sigma)]} \alpha_i(t)$$

Because eigenvalues satisfy the following indeterminate form

$$\lambda_1 > \lambda_2 > \dots > \lambda_n$$

Thus,

$$\frac{\theta_i(t)}{\theta_m(t)} < 1 \quad \forall t, i = m+1, \dots, n$$

While $t \rightarrow \infty$, $\alpha_i(t) \rightarrow 0$. For θ_m is bounded, $\theta_i(t) \rightarrow 0$ when $t \rightarrow \infty$. Derived from Eq. (22), σ is expressed as $\lambda_m \theta_m^2$, if $i=m$, Eq. (21) changes to

$$\theta_m(t+1) = [1 + \beta' \lambda_m (1 - \theta_m^2(t))] \theta_m(t)$$

So while $t \rightarrow \infty$, $\theta_m(t) \rightarrow 1$, which means the normalizing PC_m would be extracted, in the mean while, $\sigma(t) \rightarrow \lambda_m$.

Based on the analysis above, as to each neuron m ($m=1$ to N , $N \leq n$), self-adaptive PC feature extraction algorithm proceeds as follows:

Algorithm 1

- (1) Initialize p and W to small random numbers.
- (2) Select β and γ reasonably.
- (3) Update p and W according to Eq. (16), until Δp and ΔW is less then the given thresholds.

5 Experimental Analysis

As mentioned above, PCA is a useful technique to reduce the dimensionality of large data sets, such as those from microarray analysis. PCA can also be used to find signals in noisy data. The microarray data for this experiment come from. The authors used DNA microarrays to study temporal gene expression of almost all genes in *Saccharomyces cerevisiae* during the metabolic shift from fermentation to respiration. Expression levels were measured at seven time points during the diauxic shift.

In following experiment, with the help of function princomp in the Statistics Toolbox, we calculate the principal components of a data set. Princomp(X) performs PCA on the n-by-p data matrix X, and returns the principal component coefficients, also known as loadings. Rows of X correspond to observations, columns to variables.

```
[pc, zscores, pcvars] = princomp(yeastvalues)
```

pc is a p-by-p matrix, each column containing coefficients for one principal component. The columns are in order of decreasing component variance. MATLAB displays pc as follows,

```
-0.0245 -0.3033 -0.1710 -0.2831 -0.1155 0.4034 0.7887
0.0186 -0.5309 -0.3843 -0.5419 -0.2384 -0.2903 -0.3679
0.0713 -0.1970 0.2493 0.4042 -0.7452 -0.3657 0.2035
0.2254 -0.2941 0.1667 0.1705 -0.2385 0.7520 -0.4283
0.2950 -0.6422 0.1415 0.3358 0.5592 -0.2110 0.1032
0.6596 0.1788 0.5155 -0.5032 -0.0194 -0.0961 0.0667
0.6490 0.2377 -0.6689 0.2601 -0.0673 -0.0039 0.0521
```

Then, use the function cumsum to see the cumulative sum of the variances. cumsum(pcvars./sum(pcvars) * 100), detailed results as follows,

```
78.3719
89.2140
93.4357
96.0831
98.3283
99.3203
100.0000
```

This shows that almost 90% of the variance is accounted for by the first two principal components. A scatter plot of the scores of the first two principal components shows that there are two distinct regions. This is not unexpected, because the filtering process removed many of the genes with low variance or low information. These genes would have appeared in the middle of the scatter plot. Figure 1 is plotted based on data above.

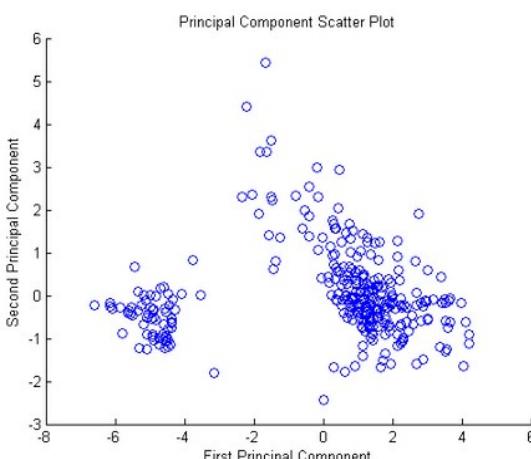


Fig. 1. Principal component scatter plot

References

1. Giles, C.L., Maxwell, T.: Learning, invariance and generalization in high order neural networks. *Applied Optics* 26, 4972–4978 (1987)
2. Cherkassky, V., et al.: Linear algebra approach to neural AM and noise performance of neural classifiers. *IEEE Trans. Computer* 40, 1429–1434 (1991)
3. Oja, E., Karhunen, J.: On stochastic approximation of the eigenvectors and eigen values of the expectation of a random matrix. *Journal Math. Analysis and Appl.* 106, 69–84 (1985)
4. Sanger, T.D.: Optimal unsupervised learning in a single layer linear feed forward neural network. *Neural Networks* 2, 459–473 (1989)
5. Kung, S.Y.: Adaptive principal component analysis via an orthogonal learning network. In: Proc. IEEE ISCAS, New Orleans (May 1990)
6. Kung, S.Y., et al.: A neural network learning algorithm for adaptive principal component extraction. In: Proc. IEEE ICASSP, Albuquerque (April 1990)
7. Kung, S.Y.: Constrained principal component analysis via an orthogonal learning network. In: Proc. IEEE ICASSP, pp. 719–722 (1990)
8. Diamantaras, K.I., Kung, S.Y.: An unsupervised neural model for oriented principal component extraction. In: Proc. IEEE ICASSP, pp. 1049–1052 (1991)
9. Xu, Y., Zhang, D., Yang, J.: A feature extraction method for use with bimodal biometrics. *Pattern Recognition* 43, 1106–1115 (2010)

Digital Library Network Based on the San Technology

Jiao Cui-Ling¹ and Wang Jian-Ping²

¹ Experiment Center, Henan Institute of Science and Technology, 453003, XinXiang, China

² School of Information Engineer, Henan Institute of Science and Technology, 453003, XinXiang, China

{jiaocuiling1969,xunji2002}@163.com

Abstract. The Application systems based on data center will be the trend of digital library network structures in the future. using the SAN technology to build the high speed digital library network will quicken the developing step of the digital library network .in this paper, we describe the core SNA technologies, structures of SAN network, and the data storage and backup systems in details.

Keywords: SAN, FCP, Data Storage, IP Storage.

1 Introduction

The rapid development of digital resources sets the new request for the higher school library network. These digital resources become the basement of the school scientific research and teaching. The high speed developments of the communication technology and the internet open up the new prospects of higher schools' information development as well as the construction of digital resources. In order to set up a sound system of digital resources, we must accelerate the digitization process of higher school library.

The following gives the basic conditions to build the digital libraries for colleges and universities: First, the internal LAN must have the high performance and the high broadband; Then, it have to introduce or purchase the advanced and can be expanded Network hardware equipment to realize the expansion of the digital library or to realize it step-by-step; Third, it must use the mass storage technology with high-speed access performance to support the growing storage need of digital information such as hypertext, graphic images, video and audio super medias and so on. Finally, it also must have the higher system security to ensure that the system will not be illegal invasion or destroyed.

The mass data storage and efficient access speed request the construction of library network storage system to adopt the SAN technology. It can absolutely meet the data storage and backup. It also provides the efficient access speed and storage speed.

SAN means Storage Area Network. It is created around Fiber Channel (FC) connections linking host computers and storage peripherals such as magnetic disk drives, tape, etc. Figure 1 provides a conceptual view of a SAN, according to this definition.

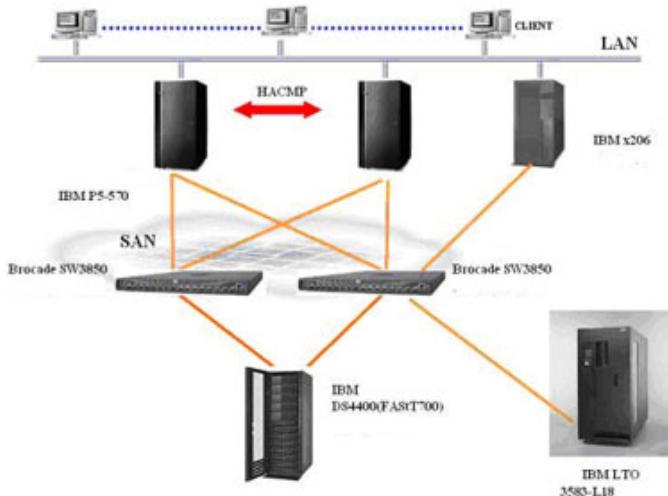


Fig. 1. The Basic Topology of SAN

Basically, some numbers of host computers are linked to a set of storage resources. How the host computers are configured or how the storage resources they are attached to are configured is immaterial. What matters is that there is an interconnection fabric that links these components together, potentially allowing any host to address any storage device that is attached to the storage network.

2 Core Technologies of SAN

The core technologies of SAN network include storage technology, backup technology and safety technology.

2.1 Storage Technology of SAN

SAN storage technology mainly includes FC-SAN and IP-SAN.

FC-SAN uses SAN network on optical fiber technology and its transmission medium is optical fiber nets. It has the high performance and is widely used. Fiber Channel is mainly used to construct the storage network with high transmission speed. Its relevant technical standards is made by the T11 technology committee of International Committee for Information Technology Standards (INCITS). Figure 2 shows the topology of FC-SAN.

In the structure FC-san usually uses optical fiber as a transmission medium. It has the feature of high transmission speed, high reliability and long distance transmission etc. At present, the transmission rate of fiber channel has reached 4Gbps, and the related products of FC-SAN whose transmission speed has been 8Gbps are also available.

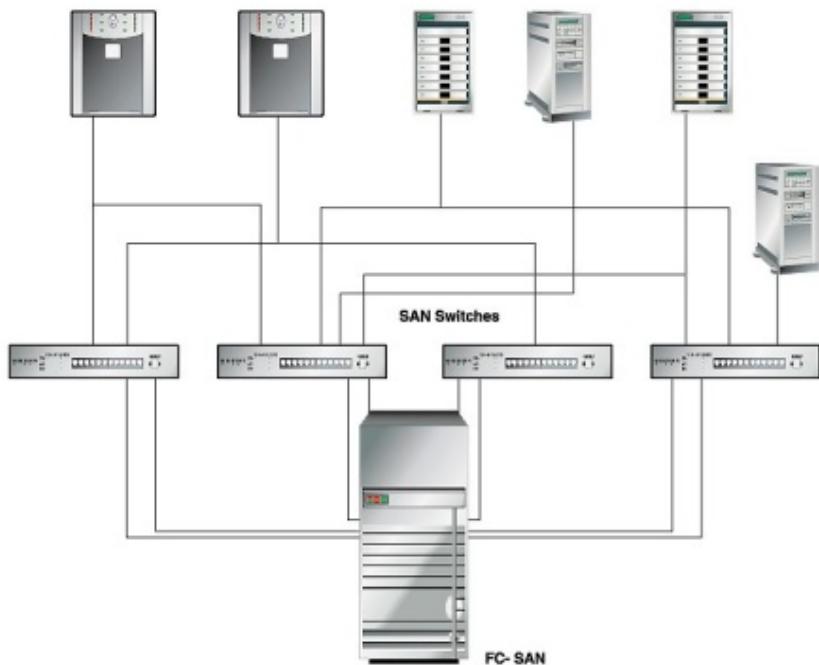


Fig. 2. Topology of FC-SAN

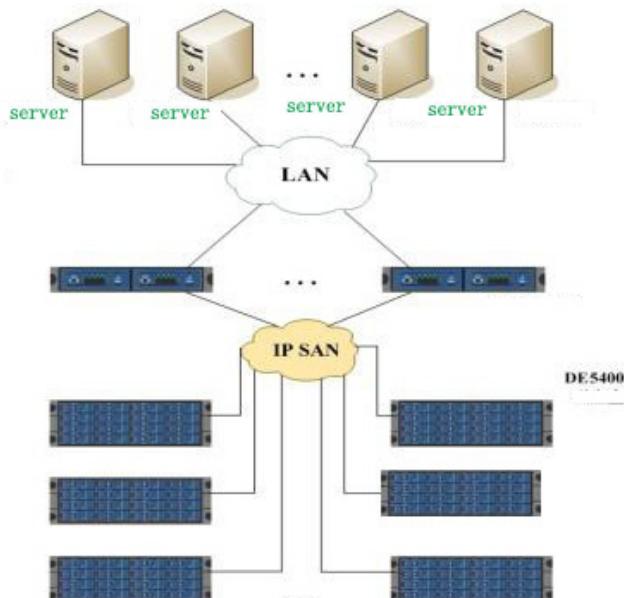


Fig. 3. Topology of IP-SAN

In FC-SAN system, the file I/O is between the application and the file-system, which talks to the volume manager, which in turn makes block I/O to the storage device. As we see here, the file-system and the volume manager (if any) are on the server and the storage has the RAID and the disks.

IP-SAN uses SAN network on iSCSI technology and its transmission medium is IP nets. IP-SAN is the storage area network constructed at the base of TCP/IP data transmission technology. Figure 3 shows the topology of IP-SAN.

It can send the SCSI instructions to the distance through the TCP communication protocol to control the remote storage device. As the transmit packet contains the IP position of transmission objectives, IP-SAN is a peer-to-peer transmission with high efficiency. The basic components of the system are the same in IP-SAN and FC-SAN, but here the application makes the file I/O to a file-system which is outside the server and the block I/O now happens within the storage between the volume manager and the RAID.

2.2 SAN Backup Technology

Data backup is one of the core tasks of storage networks. Common backup technologies of SAN are LAN Free Backup and Server Less Backup.

LAN Free Backup is the data backup technology to release network resources. In the architecture of SAN, the backup server sends instructions and information to the application server and commands the application server to backup the data from the disk array to disk base. In the process, the huge backup data flow don't flow through the network and saves precious bandwidth resources for the network.

Server Less Backup means realizing the data replication in SAN exchange layer. Then the backup data neither need through the network nor through the bus of the server and it ensures that the network and the application server operation efficiently.

2.3 Security Technologies of SAN

The safety of the SAN network relates to the high-speed network performance of the whole library. It is particularly important to establish a good SAN prevention mechanism.

2.3.1 Firewall

Firewall is a device or set of devices designed to permit or deny network transmissions based upon a set of rules and is frequently used to protect networks from unauthorized access while permitting legitimate communications to pass.

SAN network uses an enterprise firewall technology with three layer of defense circle. In outer system, it sets a good packets filtration firewall and an intrusion detection system (IDS). It uses MD5 encryption-key in the management entrance of system. To prevent the weak passwords and password brute force, it adopts the SSL identity verification to encrypt the outer core. To a lot of reading and writing data operations it needs three layer identity authentication. The firewall can configurate the

network security framework automatically to exclude the illegal connection request, eliminate illegal users and protect the internal network security of SAN.

2.3.2 Encryption Technologies

To protect the data security of SAN internal network we usually employ the tunnel encryption technology. Network senior administrator uses the assorted SAN network management software to encrypt the users' ID and password and establishes a high speed of virtual private network (VPN). Tunnel encrypted data can be effectively applied to the setting and tracking of the user privileges.

Data can be effectively applied to the setting of the user privileges and tracking. Setting exchange logic division on the SAN network. By this technology setting the storage area to three major modules includes host computer, switchboard and storage equipment. Members of the division can only communicate with other members of it, equivalent to divide the category and permission of user, the division of logic interval reducing the burden of network flow. This tunnel encrypted logic division centralizes the management of the SAN network storage equipment effectively, improving the efficiency of centered operation and monitoring.

2.3.3 Viruses Monitoring Systems

Viruses are small software programs that are designed to spread from one computer to another and to interfere with computer operation. A virus might corrupt or delete data. Viruses can be hidden in illicit software or other files or programs.

The SAN network need to interconnect with the external network, therefore, setting a virus monitoring software is very important. It is required to possess the function of heuristic virus scanning here, the function of intelligent antivirus processing, requiring the antivirus cycle is short and the antivirus data type is various. Currently speaking, the number of antivirus software product is a lot. But each has its characteristics, which can't satisfy the conditions above however. So we can choose only two sets of antivirus software as a system of the virus monitoring. The basic requirements is that the compatibility of the two types of software is better, conflicting is less, time of virus processing is short.

3 Conclusion

With the development of digital campus network construction, establishing efficient and reliable digital library by SAN arises at the historic moment, it provides more broad prospect for the construction of the digital library, storage devices in the past is directly connected to the homework on the network, SAN removed these storage device, providing more flexibility, which is convenient for equipment to upgrade, the cost is low, which forms a backup window without any restrictions. When faced with fault, the administrators of SAN network to can resume the normal work more effectively. SAN network can be located in the place away from homework nets, the construction of SAN network not only improve data storage and reading and writing skills, but enhance the data integration and unity, strengthening the data security

greatly and making it convenient for the data storage and protection. All these mean that the advantage of SAN in recent few years is even more significant. Based on the construction of the SAN digital library obviously accelerate the pace of the digital library construction.

References

1. Yin, S., Luo, Y.-Q., Zong, L., et al.: Storage area network extension over passive optical networks (S-PONS). *IEEE Communications Magazine* (46), 44–52 (2008)
2. Zhang, X., Li, Z.-H.: Simulation of IP storage area network. In: *MSV 2005*, pp. 56–59 (2005)
3. Daniel, D., Nicko, A., Vaughn, N.: Beyond the storage area network: Data intensive computing in a distributed environment. In: *Thirteenth NASA Goddard Conference on Mass Storage Systems and Technologies*, pp. 232–236 (2005)
4. He, X.-B., Zhang, M., Yang, Q.: STICS: SCSI-to-IP cache for storage area networks. *Journal of Parallel and Distributed Computing* (64), 1069–1085 (2004)
5. Gauger, C.M., Köhn, M., Gunreben, S.: Modeling and performance evaluation of iSCSI storage area networks over TCP/IP-based MAN and WAN networks. In: *Broadnets*, pp. 915–923 (2005)
6. Sunil, G., Francois, B., Eric, F.: Pb-free process development for a high end storage area network application. In: *Proceedings of the ASME InterPack Conference*, pp. 937–943 (2007)
7. Ehrhardt, A.: Extension of storage area networks and integration on different platforms of an optical transport network. In: *International Conference on Transparent Optical Networks*, pp. 229–232 (2004)
8. Lin, S.-B., Lu, M.-H., Chiueh, T.-C.: Transparent reliable multicast for Ethernet based storage area networks. In: *International Symposium on Network Computing and Applications*, pp. 87–94 (2007)
9. Bursky: Security processor targets IP storage networks Dave. *Electronic Engineering Time*, 32–33 (2006)
10. Wu, F., Xie, C.-S., Hu, D.-Q.: The design and research of high-speed channel for volume holographic data storage. In: *The International Society for Optical Engineering*, pp. 1072–1076 (2004)
11. Andrea, B., Paolo, G., Maria, G.E.: Multicast support for a storage area network switch Globecom. In: *IEEE Global Telecommunications Conference*, pp. 1040–1047 (2007)

Websites Based on J2EE of Struts

Jiao Cui-Ling¹ and Wang Jian-Ping²

¹ Experiment Center, Henan Institute of Science and Technology, 453003, XinXiang, China

² School of Information Engineer, Henan Institute of Science and Technology, 453003, XinXiang, China

{jiaocuiling1969,xunji2002}@163.com

Abstract. Establishing Web service based on the layer structure could raise the efficiency of Web development, it could manage the system structure effectively; lower the complications and the supporting difficulty. The important technology of MVC software designing model was the separating of the operation code with the view code, struts was the implementation of the MVC model based on the Web form according to the J2EE. It was analyzed the MVC software designing model, JSP MODEL 2 frames, the main technology of the struts, expatiated the advantage and disadvantage of the Web development form based on the struts in details.

Keywords: Struts, MVC, J2EE, JAVA.

1 Introduction

Java is used in a wide variety of computing platforms from embedded devices and mobile phones on the low end, to enterprise servers and supercomputers on the high end. The development platforms of java can be divided into Java SE, Java EE and J2ME.

J2SE is used for general-purpose on desktop PCs, servers and similar devices. J2ME is the Micro Edition of JAVA. It specifies several different sets of libraries for devices with limited storage, display, and power capacities. It often used to develop applications for mobile devices, PDAs, TV set-top boxes, and printers.

J2EE means Java Platform, Enterprise Edition or Java EE. It is Oracle's enterprise java computing platform. The platform provides an API and runtime environment for developing and running enterprise software, including network and web services, and other large-scale, multi-tiered, scalable, reliable, and secure network applications. J2EE defines the core set of API and features of Java Application Servers. The Web modules include servlets, and JavaServer Pages.

JSP (JavaServer Pages) is a Java technology that helps software developers serve dynamically generated web pages based on HTML, XML, or other document types. Architecturally, JSP may be viewed as a high-level abstraction of Java servlets. JSPs are loaded in the server and are operated from a structured special installed Java server packet called a Java EE Web Application, often packaged as a file archive.

With the gradual development and wide application of the J2EE system, some excellent application frames appeared. Among these frames, Struts was a more

popular one. It was a project which ASF supported, by using the technology of JSP ,it accomplished the application frames based on the MVC design form of J2EE web, Struts was a categorical product of MVC, It made the developer free from the trivial work and exploited further web application ,Saving the time of designing frames, struts let the developer concentrate on how to cope with the problem of real business ,which greatly simplified the development process of web application .Thus it shortened the development period and improved the development efficiency.

2 MVC Model

MVC is short for model view controller. MVC is a software architecture, currently considered an architectural pattern used in software engineering. The model manages the behavior and data of the application domain, responds to requests for information about its state, and responds to instructions to change state. In event-driven systems, the model notifies observers when the information changes so that they can react. MVC was a design form and it compelled to part the input, treatment and output of the application program. It divided the application program into three core modules: pattern view and controller, which undertook different task.

Model: business logic layer, realizing the function of specific business logic and state control.

View: expression layer, reading the interface of interaction with the user and the function of date input and output. The view renders the model into a form suitable for interaction, typically a user interface element. Multiple views can exist for a single model for different purposes. A view port typically has a one to one correspondence with a display surface and knows how to render to it.

Controller: control layer .Realizing the function of Flow control and the cooperation between view and Model. The controller receives user input and initiates a response by making calls on model objects. A controller accepts input from the user and instructs the model and a view port to perform actions based on that input.

The relation of Model, View and Controller shows as figure 1.

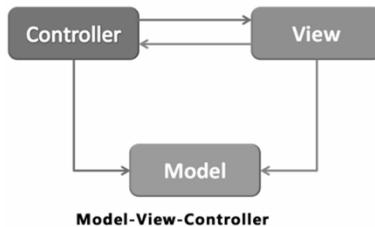


Fig. 1. Relation of MVC

3 JSP Models

In the design of Java Web applications, there are two commonly used design models, referred to as Model 1 and Model 2.

3.1 JSP Model 1

In JSP Model 1, a request is made to a JSP or servlet and then that JSP or servlet handles all responsibilities for the request, including processing the request, validating data, handling the business logic, and generating a response. The Model 1 architecture is commonly used in smaller, simple task applications due to its ease of development. The Structure of JSP Model 1 shows as figure 2.

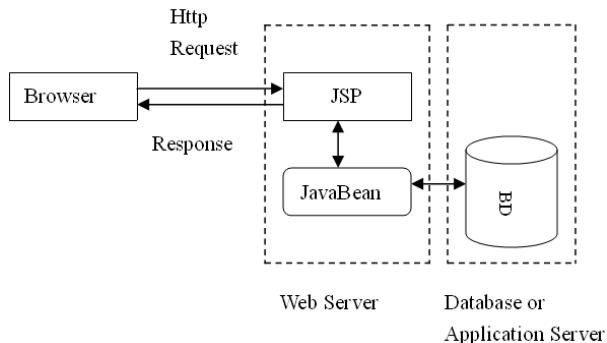


Fig. 2. JSP Model 1 Structure

3.2 JSP Model 2

JSP Model 2 is a more complex design pattern that separates the display of content from the logic used to obtain and manipulate the content. In a Model 2 application, requests from the client browser are passed to the controller. The controller performs any logic necessary to obtain the correct content for display. It then places the content in the request (commonly in the form of a JavaBean or POJO) and decides which view it will pass the request to. The view then renders the content passed by the controller. The Structure of JSP Model 2 shows as figure 3.

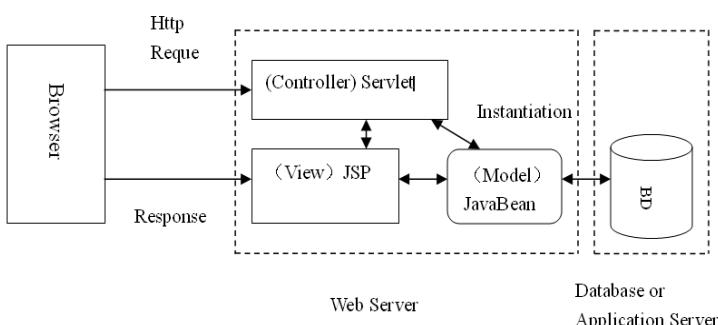


Fig. 3. JSP Model 2 System Structure

4 The Core Technology of Struts

4.1 The Core Module of Struts

(1) ActionServlet module

Struts' controller is mapping the events (the events usually are HTTP post)to a Servlet of the module. ActionServlet is the realization of the MVC Command, and it is the core of this framework. ActionServlet creates and uses Action, ActionForm and ActionForward. By modify the file struts-config.xml to configure the command. In creat Web project, the user through extending Action and Actionform to solve the specific problem. File struts-config.xml instructs Actionservlet that how to use the expanded module.

(2) ActionForm module

ActionForm maintains Web applications session state. ActionForm is an abstract module, and it must be for each input form type of the module to create more children. ActionForm said is set by HTML form or updated the general sense of the data. In setting UserActionForm form of organization by HTML,Struts framework will examine whether there is UserActionForm, if does not exist, it will create an instance of the module. By use the corresponding domain HttpServletRequest to set up the state of UserActionForm. Struts framework will be passed to the User ActionForm in business and will be updated it before packing the state of User Action. Before passing it on to the kind of Action,Struts will also examine UserActionForm the form state by reallocating the method of validation().

(3) Action module

Action module is a package of the logic business. The purpose of the module makes HttpServletRequest convert to logic business. To use Action, you should create submodule of it and cover process() method. By using perform()method, ActionServlet(Command)will pass the parametric module to ActionForm.

In the Struts,Action is a guide to action, and passes the available information to the corresponding logic modules to handle. Servlet is inherited by Action, so that some simple function can be placed in Action, at last the Function perform()returns operation success or failure of state records.

(4) Error module

The defined Error categories of Struts include ActionError and Action-Erroes. ActionError encapsulates the single error messages. ActionErroes is the container of the ActionError, and View can make the popularization and application of access to these markers. ActionError is made that the Struts to keep way of the error list.

(5) ActionMapping module

ActionMapping includes how the given incident mapped to the given Action information. Through the way of perform(), ActionServlet (command) passed the ActionMapping to Action module. In this way it allowed Action to visit the information used for control flow.

4.2 Struts Work Flow

In the Web application process of Struts Frames, it would load and initialize ActionServlet when Web application was started. ActionServlet read pre-load information from the document of struts-config.xml and put them in the respective disposal target. When ActionServlet receives a user's demand, it would implement the following flow: judge whether the suffix of a user's HTTP demand is "do" and all the demand would be intercepted by control ActionServlet. According to the collocating information in Struts-config.xml, ActionServlet maps the intercepting information to the Action module accordingly, and sealed the user request form in FormBean with a given name. Then ActionServlet called the way of execute() in business logic Action case to implement business logic. After using this method, it would return a ActionForward target. Controller would search the collocating information based on the returned target and turn the JSP page accordingly.

5 Conclusion

The advantages of using Struts to realize the website construction: It adopts JSP marked mechanism to get the reusing codes and abstract codes. This method contributed to improve the reusability and flexibility of codes. When the technology space or the problem space varied, users have much more chances to reuse codes. Using open source, everybody in the user's room could check up codes. Struts make the designers and developers pay their attention to their own favorite aspects. Adopting separately control ideology to manage the problem space. The websites based on this pattern had perfectly robustness. The layer structures were clear. As to huge scale of system, Struts conducted to manage complicated system.

The disadvantages of using Struts to realize the website construction: The applicable scope of Struts is limited. Struts are MVC solution based on web. So it must be achieved by HTML, JSP document and Servlet and use J2EE application program to support Struts. Struts need to support JSP1.1 and Servlet 2.2 standard Servlet container. Of course, isolating the problem scope, but strengthening the complicity, so one must accept some training before they adopt Struts development.

References

1. Selfa, D.M., Carrillo, M., Del Rocío Boone, M.: A database and web application based on MVC architecture. In: International Conference on Electronics, Communications and Computers, pp. 48–49 (2006)
2. Lin, Y.-L., Hu, J.-P.: Design and implementation of the internet service platform for rural house building technique criteria and management based on .NET MVC. Applied Mechanics and Materials, 1727–1731 (2011)
3. Wang, L.-H., Xi, M., Li, D.-X.: A network-friendly architecture for Multi-view Video Coding (MVC). Advanced Materials Research, 678–681 (2010)
4. Li, J.-M., Ma, G.-S., Feng, G., Ma, Y.-Q.: Research on Web Application of Struts Framework Based on MVC Pattern. In: Shen, H.T., Li, J., Li, M., Ni, J., Wang, W. (eds.) APWeb Workshops 2006. LNCS, vol. 3842, pp. 1029–1032. Springer, Heidelberg (2006)

5. Xu, C.-W.: Applying the MVC diagram for building web applications in struts. In: Proceedings of the 2005 International Conference on Internet Computing, pp. 38–44 (2005)
6. Li, Y., Chen, Q.-J.: MVC mode-based mobile telerobotic system and its java implementation. In: Proceedings of the World Congress on Intelligent Control and Automation (WCICA), pp. 4736–4740 (2004)
7. Zhang, Z.-Q., Xu, C.-J.: MVC model based on high-performance computing operating audit system. In: International Conference on MultiMedia and Information Technology, pp. 241–243 (2010)
8. Zhang, H.-T., Duan, F.-J., Wang, X.-Y.: Application framework of embedded system based on MVC design pattern. Journal of Harbin Institute of Technology, 166–168 (2009)
9. Chen, N.: Automatic generation of resource access code based on MVC mode. Journal of Shenyang University of Technology, 417–422 (2010)
10. Cui, W., Huang, L., Liang, L.-J.: The research of PHP development framework based on MVC pattern. In: International Conference on Computer Sciences and Convergence7. Information Technology, pp. 947–949 (2009)

Research of a Vertical Search Engine for Campus Network

Rujia Gao, Wanlong Li, Shanhong Zheng, and Hang Li

College of Computer Science and Engineering Changchun University of Technology
Yan'an Road 2055 Changchun City Jilin Province
gaorujia723@126.com

Abstract. This paper employs the campus network as an experimental platform, using the characteristics of the vertical search engine and the system architecture of Lucene, designing and building a vertical search engine system for campus network in web crawler, indexing, searching and interface design. The vertical search engine is a specialized search engine focus on some industry-specific and some specific populations for collecting data and extracting information of the structure of HTML, it's a subdivision and extension of search engine, it can provide more accurate and more professional search results for the users.

Keywords: vertical search engine, lucene, web crawler, indexing, searching.

1 Introduction

With the rapid expansions of the Internet and the software and hardware technology, the Web information is also explosive growth. CNNIC put forward that: “the 29th China Internet Development Statistics Report” has been released on January 16, 2012. By the end of 2011, the total number of websites in China achieved 86.6 billion, increasing by 44.3 percent more than the same period in 2010; the scale of Chinese websites reached to 2.296 million, growing by 20 percent by the end of 2010; the scale of China’s cyber citizen up to 513 million, the Internet penetration rate rose to 38.3 percent, up four percentage more than the same period of 2010, increasing by 55.8 million Internet population [1]. Compared to China, the total number of web pages will be more abundant. It has been found that search engine is known as the commanding heights of the Internet. The traditional search engine that is general search engine, it can return more search results to Internet users, it is difficult to look up accurate results and relevant information at the same time of mass information, which leads to a lot of redundant information and spam, it’s too hard to find information that is needed for the Internet users. Then, in order to meet the requirements of accuracy and speed for the Internet users to retrieve information, a more professional, more accurate and more targeted search engine occurred, that is vertical search engine.

This paper put forward the construction and research of a vertical search engine based on campus network, through this method, the Internet users can easily and quickly set up search engine based on Lucene and Windows workbench in MyEclipse development. The paper's work to be done is to achieve Key words semantic search in the latter. I will improve it to make it more comprehensive and more accurate on search keywords in the latter.

2 Vertical Search Engine and Lucene

2.1 Research and System Structure of the Vertical Search Engine

The vertical search engine is a specialized search engine focus on some industry-specific and some specific populations for collecting data and extracting information of the structure of HTML, it's a subdivision and extension of search engine, it can integrate a kind of special information about webpage and extract the unstructured data into structured data by the way of directional branch field, it can return search results to the Internet users by unstructured and structured way after processing. Currently the direction of vertical search engine is quite a lot, such as shopping search, talent search, MP3 search and property search and so on, almost various walks of life all kinds of information can become a respective vertical search engine. As is shown in Figure 1:

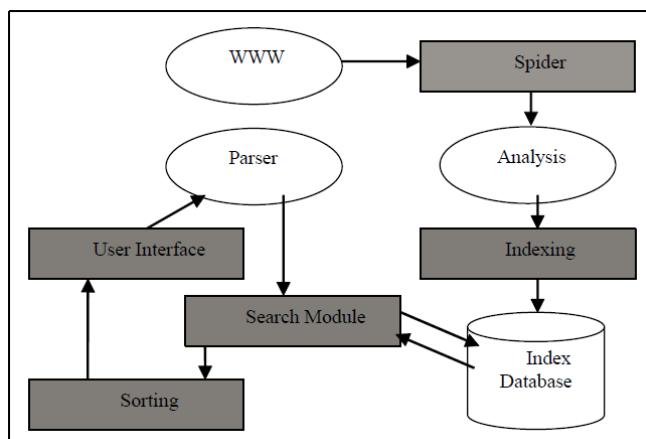


Fig. 1. System Structure of Search Engine

2.2 Overview and System Structure of Lucene

Lucene is a subproject of the Apache Software Foundation Jakarta project [9]; Lucene is a full-text search engine framework rather than a product. So it's unlike Baidu or

Google as usual we use rather to provide users with a tool to achieve these products. Lucene is a set of Java API, it is not a unique search engine system, but you can use it to develop the search engine system. The retrieve object of Lucene is colorful, such as xml, documentation, database and so on. These can be used as its retrieve object.

Lucene is an excellent full-text search engine; its system structure has a strong object-oriented characteristic. There are seven modules in the Lucene source code package; the five main modules are shown in Table 1.

Table 1. The Function table of Lucene Source Code Package

org.apache.lucene.analysis	Analyzer, it's used to segment the document.
org.apache.lucene.document	Document Management, it mainly responsible for the management of fields.
org.apache.lucene.index	Indexing management, it's used to insert records, delete records and so on.
org.apache.lucene.search	Retrieval Management, according to the query to obtain the results of retrieval.
org.apache.lucene.queryParser	Query Parser, parsing the user's query and passing to searcher.

3 The Construction and Research of Search Engine

To build a campus network of a vertical search engine is mainly composed of four parts: web crawler, document parsing, indexing storage, retrieval output and interface design, one of the most crucial and most core module is indexing storage and retrieval output. Firstly, this paper uses web spider to crawl the school website and secondly, it uses regular expressions to parse the document and filter the information that are not used; finally creates Web application in My-Eclipse for users to test in the browser address bar.

3.1 Crawl Module

Web crawler is also known as web spider, it is a program which sent a get request to the site, it can get first page from the school website once in a while (10 minutes), if it found a new link then put it to the search results, then follow the link to continue to access a new page, when it meets a certain condition, it will automatically stop crawling. Firstly, this paper use the Web crawler to crawl information needed by the user in the web on the network, deposited in local library according to the screening; then using the full-text search tool to screen the data which is crawled; finally, use of the weights of a series of algorithms and sort out the screening results presents to the user. Web spider will not crawl JavaScript, The search engine can not see page through JavaScript to dynamically generate.

3.2 Analysis Module

Analysis is as implied by the name; it will divide the full-text into multiple terms. The term can be retrieved by the user and then put the term on the index database when creating index. Term is also called text which is processed and segmented by the analyzer. Only through analyzing can make information retrieval system understand the user's retrieval request, so as to search for related information. The word segmentation which is used to indexing and searching is the same shown in Figure 2:

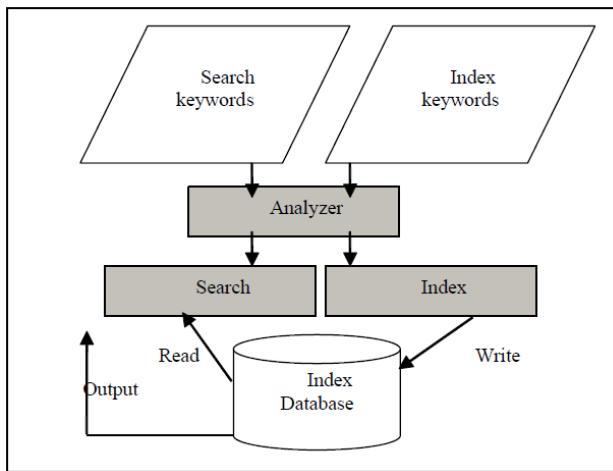


Fig. 2. Background module of search engine

English words are separated by spaces, it is almost impossible to express correct meaning and accurate word segmentation for Chinese word segmentation. Chinese word segmentation is mainly composed of the following three ways: (1) Word segmentation divides the full-text into single word when analyzing the Chinese texts. For example “we are good students”, after analyzing the words is “we”, “are”, “good”, “student”. (2) Dichotomy creates indexing and uses each of two words as a term. For example “we are good students”, the result is “we are”, “are good”, “good students”. (3) Dictionary divides the encountered text into words through constructing a commonly dictionary and uses it to match word in the text, such as the maximum forward/backward matching algorithm and the least forward/backward matching algorithm. From the results, the last one is more precise, in order to improve the accuracy of segmentation is the best combination of two used together due to the Chinese word segmentation of regeneration and expansion.

3.3 Indexing Module

It is a duty to understand and analyze the information which crawled by the web crawler for the indexer, extracting index entry from the relevant information and

using it to represent the documents and generating the index table of the document library and creating index database at the same time. There are two kinds of indexing in the index entry of Lucene, they respectively are objective indexing and content indexing. The former one is to collect some objective information about the documents and has nothing with the semantic content of the document, such as author name, update time, the encoding and so on; the latter one has something with the semantic content with the documents, such as keywords, statements and weights and so on. The system is based on the way of inverted index, whether the indexing algorithm is good or bad has an effect on the indexer response time. The index part of the code is as follows:

The system employs MVC pattern to create an index which the information the spider crawling down from the school, if the length of the title and URL address is one, you can not index these; if the title and URL are both occur, we can add the field of title, URL and content into the Document of Lucene, then add the Document to the index of Lucene and finally closing the indexer.

3.4 Search Module

The searcher's main task is to receive and interpret the user's query request, it based on keywords that the user input to quickly retrieve the document which the user needs in the index library, evaluate the documents and keywords, sort the information that is relevant with the Internet users according to the evaluation results, and realize relevance feedback of the user. The search part of the code is as follows:

```
MamboSearchResult search (String search Word, String con,
int pagenum, int contentnum )

Searcher searcher = create Searcher ()

Query Parser parser = new Query Parser
(Version.LUCENE_30, "content",new
StandardAnalyzer(Version.LUCENE_30))

String queryPattern =search Word

TopScoreDocCollector collector =
TopScoreDocCollector.create (100000, false)

Searcher. Search (query, collector)

ScoreDoc[] hits = collector.topDocs().scoreDocs

reHtml=cn.display(totalRows,result.getResults(),
searchWord,10,pageNumber)

showPageHtml=cn.showPage(totalRows,pageNumber,10,
searchWord, "")
```

3.5 Interface Design

The user interfaces of this system uses JSP technology to develop, taking Tomcat as the Web server, including four main interfaces: they are separately index.jsp, IndexCreating.jsp, search.jsp and searchresult.jsp. Search page is used for receiving the keywords (multi-keyword search) which the user inputs it can be displayed the information that matches the search keywords to the users and other associated word will be displayed in the interface. Index page is used to create an index of the text the spider crawls. This system mainly uses the pattern of MVC to implement, at the present this system can only design the user interface by indexing module and searching module of vertical search engine based on campus network, but in the future I will add the semantics of the keywords and Chinese word segmentation to my vertical search engine so as to make it more perfect as much as possible.

Development tools used in the development process of this system: apache-tomcat-6.0.33, jdk1.6.0_07 and My Eclipse 8.5. When the user tested this system, we need to start Tomcat server at first, and then input the browser address: <http://localhost:8080/search/index.jsp>, jump to the indexing module. If you input <http://localhost:8080/search/searching.do> in the address bar can enter into the search interface.

At present the system can show amount of information to the user that is relevant to search keywords in the result page and can highlight the search keywords, can only displays one piece of information, can be capable of extracting the first 50 words to display. The next job I am going to do is to display semantic information which is related to relevant and similar search keywords. Or search engine searches can be based on keywords user inputs to understand its meaning. This requires the knowledge of ontology, due to the rush of time, part of the contents in this system is not well completed, and I will continue further research in the future. The searcher's main task is to receive and interpret the user's query request, it based on keywords that the user.

4 Conclusions

This paper mainly sketches the design and implementation of vertical search engine based on campus network combined with Lucene, analyzing and researching the main modules. Regarding the school campus network as an experimental platform, combined with relevant knowledge of Lucene to build certain professional and certain field search engine system based on campus network. Since Lucene is a full-text search engine toolkit which is free and open source, so we can modify some algorithm. However, the system still exist some flaws, such as the semantics of the keywords and Chinese word segmentation problem; the above problems are needed to further study.

References

1. China Internet Network Information Center. The 29th China Internet Development Statistics Report (EB/OL), http://www.cnnic.net.cn/dtgg/dtgg/201201/t20120116_23667.html

2. Cai, J., Guo, Y., Wang, L.: Design and Implementation of School Search Engine Based on Lucene. *Net. Computer Technology and Development* 16(11), 73–75 (2006)
3. Zhou, Y.: Design and Implementation of a small search engine system based on Lucene. *Networking and Communications* 12(10), 61–64 (2010)
4. Lan, Y.: The Design and Implement of a Vertical Search Engine for Campus Network Based on Lucene. *Guangdong Education Institute* 29(3), 98–103 (2009)
5. Si, M.: Construction and Research of school website based on vertical search engine. *Technology Online* 21(01), 108–109 (2011)
6. Li, H., Ye, F., Bian, L., et al.: The Research and Implement of Education Network Vertical Search Engine. *Network and Communication* 26(10-3), 108–110 (2010)
7. Zhao, Q., et al.: Information Retrieval. China Machine Press, Beijing (2008) (in Chinese)
8. Qiu, Z., Fu, T.: Develop its own search engine Lucene 2.0 + Heritrix. People Post Press, Beijing (2010) (in Chinese)
9. Che, D.: In the application to add full-text search function-Lucene full-text indexing engine based on Java (EB/OL), <http://www.chedong.com/tech/html>
10. Gospodnetic, O., Hatcher, E.: Lucene in action. Maning Publications Co. (2005) (in Chinese)
11. Li, Y., Ding, H.: Research and Application of Full-Text Search Based on Lucene. *Computer Technology and Development* 20(02), 12–15 (2010)
12. Geng, X., Zhang, Y.: JSP Essentials. Qinghua University Publications, Beijing (2009)

Efficient Control Scheme for Surface Temperature of Hot Roller Based on Neural Network

XueLi Wu, Lifeng Hou, and Hui An

Hebei University of Science and Technology, Shijiazhuang 050018, China
wuxueli@yahoo.com.cn, hlf123hlf@yeah.net, anhui96012@126.com

Abstract. Hot roller as important heating equipment widely used in textile industry, it mainly used as drying device or heating device, the surface temperature of hot roller needs to be strictly controlled. The temperature variation process has its own characteristics. This paper studies the application of neural network control in temperature control of hot roller. Neural network control can achieve good control performance in temperature control. This paper also introduces how to use distributed control system to control hot rollers with neural network control.

Keywords: neural network control, hot roller, temperature.

1 Introduction

Hot roller is widely used in textile industry, such as used for drying fabrics or heating fiber. At present, there are many hot rollers used in textile industry. A complete set of hot roller consist of heating part and drive part. The drive part keeps hot roller rotation. The heating part controls the roller's temperature. Heating part is the most important part in the hot roller, and temperature control system is core section of heating part. Considering the heating process has strictly demand for temperature precision. Hot roller heating process has following characteristics: big capacity inertia, time-delay. So the research for hot roller heating control is very important.

At present, the control system of hot roller mainly adopts to traditional PID control system. This method has many advantages, such as simple structure, highly control precision, less parameters, and parameter setting simple etc. Therefore traditional control method has widespread application in engineering control field. But this control method based on accurate mathematical model, it is difficult to obtain satisfactory control effect on the basis of the actual situation. In addition, the traditional control method needs long time to set parameters and hard to get the real global optimal value, so it is difficult satisfy the control system dynamic quality and the steady precision at the same time. In order to overcome the weakness of the traditional control method, researchers propose many improve project for traditional control. In recent years neural network get much attention. Neural network has following characteristics: it is not dependent on mathematical model, information distributed storage, parallel processing ability is strong, and it has self-study ability.

According to the characteristics of the hot rolling heating process, this paper combines neural network with traditional control method, using neural network PID control method to control heating process of hot roller in order to get better control performance. This paper includes the research of the new method, and one realization scheme based on DCS system.

2 Neural Network Controller Design

This section introduces traditional control method first. Then propose a new method which combines NNC with traditional control method. Using new control method can improve control performance of heating process.

2.1 Digital PID Control

PID controller is widely used in industrial production, its control law can express as follow:

$$u(t) = k_p e(t) + k_i \int_0^t e(t) dt + k_d \frac{de(t)}{dt} \quad (1)$$

Through the discrete can get digital increment PID controller expression.

$$\begin{aligned} u(k) = & u(k-1) + k_p [e(k) - e(k-1)] + k_i e(k) \\ & + k_d [e(k) - 2e(k-1) + e(k-2)] \end{aligned} \quad (2)$$

k is sampling period, u is control output, $e(t)$ is error signal, k_p is proportionality factor, k_i is integral coefficients k_d is differential factor.

Traditional PID control method based on accurate mathematical model, but the actual industrial production process usually hard to get accurate mathematical model. Because of the process is nonlinear and has large time delay, the traditional digital PID control is often very difficult to get satisfactory control effect. This paper introduce neural network to solve the problem.

2.2 Neural Network Control Design

$x_i(k)$ is neurons input, $\Delta u(k)$ is neurons output, K is proportionality factor, $K > 0$, w_i ($i=1,2,3$) is weighted value.

$$y_r(k) = (1-\alpha)r(k-1) + \alpha y_r(k) \quad (3)$$

By selecting $\alpha (0 < \alpha < 1)$ can get better control effect and make the system get good robustness.

According to the digital PID algorithm structure, select input as:

$$\begin{cases} x_1(k) = e(k); \\ x_2(k) = e(k) - e(k-1); \\ x_3(k) = e(k) - 2e(k-1) + e(k-2) \end{cases} \quad (4)$$

So the neural network PID controller output can express as:

$$\Delta u(k) = K \sum_{i=1}^3 w_i(k) x_i(k) / \sum_{i=1}^3 |w_i(k)| \quad (5)$$

By adjusting the weight coefficients, neural network control can realize self-adaptive. Hebb learning is the most widely used method now, use this learning rule to adjust the weighted coefficient.

$$u(k) = u(k-1) + K \sum_{i=1}^3 w_i(k) x_i(k) \quad (6)$$

$$w_i(k) = w_i / \sum_{i=1}^3 |w_i(k)| \quad (7)$$

$$\begin{cases} w_1 = w_1(k-1) + \eta_p z(k) u(k) x_1(k) \\ w_2 = w_2(k-1) + \eta_i z(k) u(k) x_2(k) \\ w_3 = w_3(k-1) + \eta_d z(k) u(k) x_3(k) \end{cases} \quad (8)$$

Where $z(k) = r(k) - y(k) = e(k)$, and η_p, η_i, η_d are corresponding learning rate.

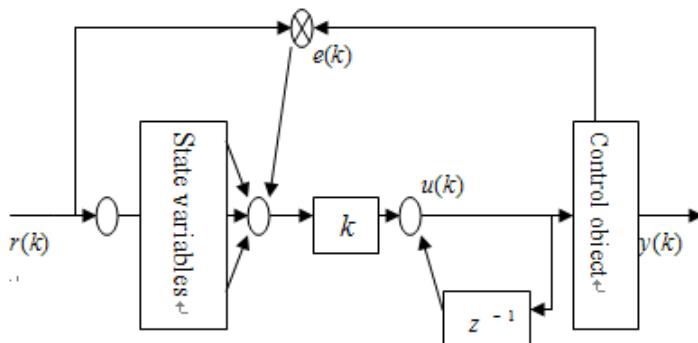


Fig. 1. Control system diagram

In the neural network PID control, integral coefficient, proportional factor, differential factor usually adopt different learning rates, in order to adjust weight according to their respective needs.

From structure diagram, it is obvious that w_1, w_2, w_3 corresponding to k_p, k_i, k_d . By adjusting neuronal structure weight, neural network PID control can implement online learning function. The control effect depends on learning algorithm of weight and reasonable choice of learning rate.

In addition, the choice of neuron proportion coefficient K is very important. When K is big, system has a good rapidity property, but the overshoot of system is big. When object has large time delay, the K value needs to be reduced in order to ensure the stability of the system. But when the K too small, the system has a slow rapidity.

3 Control System Design

3.1 Hardware Introduction

Hot roller always not used alone in the practical application. Hot rollers that in the same equipment need to cooperate with each other. So requires coordinate control.

Distributed Control System (DCS) is widely used in industrial field. DCS has strong ability of data processing, and it can solve the problem of multiple hot rollers collaborative work.

DCS mainly consist of field workstation and engineer station etc. External signal acquisition and output control signal are realized by field workstation. Engineer station receives the data from field workstation. Then engineer station process the data and output control signal. The control algorithm stored in the engineer station.

This paper takes single hot roller control system design for example.

Flowchart as follow:

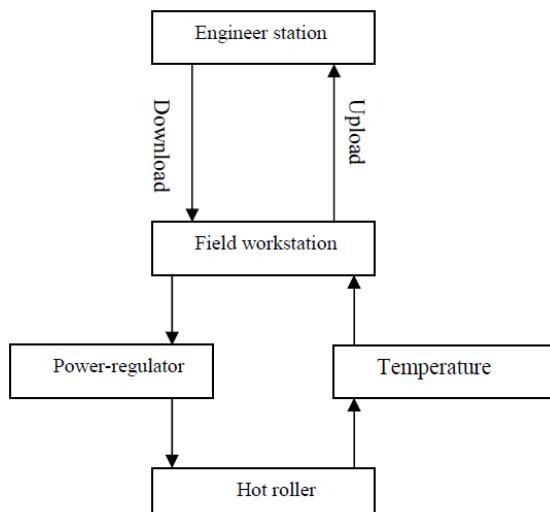


Fig. 2. Work flow chart

Field workstation collects mainly get temperature data from hot roller through temperature sensors. Then the data is delivered to engineer station. After data processing engineer station transmit control to the field workstation. Field workstation control power-regulator power output in order to adjust the temperature of hot roller.

3.2 Software Design

DCS can implement a variety of control algorithm by programming different procedure. Neural network PID control algorithm also can be implemented by appropriate procedure.

Program flow char as Fig.3.

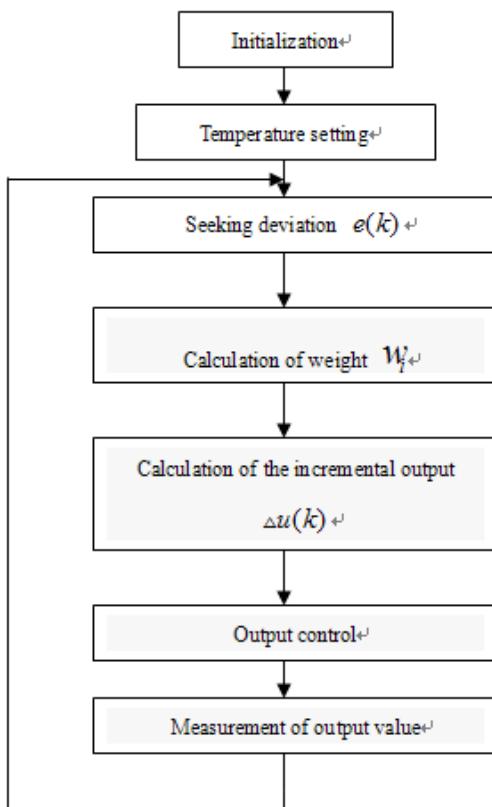


Fig. 3. Program flowchart

4 Simulation

Suppose the mathematical model of hot roller is:

$$G_p(s) = \frac{5}{20s+1} e^{-20s}$$

Control model contains lag link. Inputting unit step signal, and select learning rate as:

$$\eta_p = 0.4; \eta_i = 0.35; \eta_d = 0.4$$

Simulate it with Matlab, and compare with traditional PID controller.

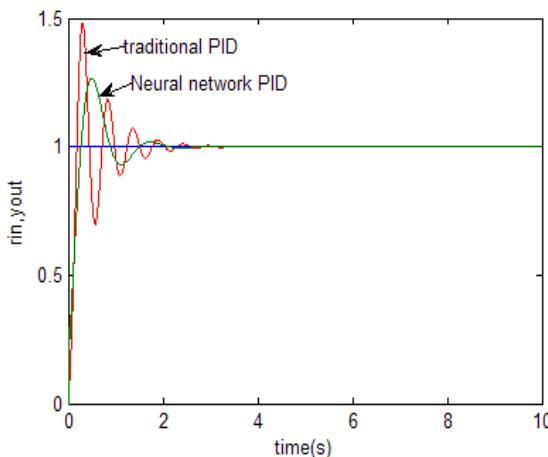


Fig. 4. State response

Compare neural network control method with traditional control method, it is obvious that neural network control method reduces overshoot significantly and improves the response speed of system. Through the simulation can prove that this method can achieve good control performance when the system contains large time delay or the mathematical model of system is not accurate.

References

1. Shen, Y.: Adaptive online state-of-charge determination based on neuro-controller and neural network. *Energy Conversion and Management* 51(5), 1093–1098 (2010)
2. Liu, L., Luo, J.: Research of PID Control Algorithm Based on Neural Network. *Energy Procedia* 13, 6988–6993 (2011)
3. Chen, J., Huang, T.-C.: Applying neural networks to on-line updated PID controllers for nonlinear process control. *Journal of Process Control* 14(2), 211–230 (2004)
4. Kansha, Y., Jia, L., Chiu, M.-S.: Self-tuning PID controllers based on the Lyapunov approach. *Chemical Engineering Science* 63(10), 2732–2740 (2008)

5. Lee, C.-H.: Calculation of PID controller parameters by using a fuzzy neural network. *ISA Transactions* 42(3), 391–400 (2003)
6. Zaheer-uddin, M.: Neuro-PID tracking control of a discharge air temperature system. *Energy Conversion and Management* 45(15-16), 2405–2415 (2004)
7. Lima, J.M.G., Ruano, A.E.: Neuro-genetic PID autotuning: time invariant case. *Mathematics and Computers in Simulation* 51(3-4), 287–300 (2000)
8. Zhang, Y.: Modeling and Control of Nonlinear Discrete-time Systems Based on Compound Neural Networks. *Chinese Journal of Chemical Engineering* 17(3), 454–459 (2009)
9. Wu, X.: Neural network-based adaptive tracking control for nonlinearly parameterized systems with unknown input nonlinearities. *Neurocomputing* 82, 127–142 (2012)
10. Shu, H., Pi, Y.: PID neural networks for time-delay systems. *Computers & Chemical Engineering* 24(2-7), 859–862 (2000)
11. Bhullar, R.S.: Strategies for implementing advanced process controls in a distributed control system (DCS). *ISA Transactions* 32(2), 147–156 (1993)
12. Chan, K.C., Leong, S.S., Lin, G.C.I.: A neural network PI controller tuner. *Artificial Intelligence in Engineering* 9(3), 167–176 (1995)
13. Carvajal, J., Chen, G., Ogmén, H.: Fuzzy PID controller: Design, performance evaluation, and stability analysis. *Information Sciences* 123(3-4), 249–270 (2000)
14. Ng, G.W., Cook, P.A.: Real-time control of systems with unknown and varying time-delays, using neural networks. *Engineering Applications of Artificial Intelligence* 11(3), 401–409 (1998)
15. Camacho, E.F., Alaré, M.R.: Neural network based adaptive control. *Annual Review in Automatic Programming* 19, 13–24 (1994)

Online Shopping System Based on Improved MD5

Ji-Rui Li¹ and Su-Qing Zhang²

¹ Information Engineer Department, Henan Vocational and Technical Institute Lecturer,
Zhengzhou, China

² Information Engineer Department, Henan Vocational and Technical Institute Associate
Professor, Zhengzhou, China

Abstract. Message-Digest Algorithm 5, a hash function which is widely used in the field of computer security, can provide the message integrity protection. But to the field following certain rules, for example, by date of birth as a password, using the MD5 encryption algorithm, can still quickly decipher by MD5 decoding procedures of list form. In view of this situation, on the basis of MD5, putting forward a improved algorithm which has greatly security in data safety, and the improved algorithm is applied to the online shopping system based on ajax framework, for digital signature function of users on the system provides powerful safety security.

Keywords: MD5, Encryption Algorithm, Information Security.

1 Encryption Algorithm

Encryption algorithm is usually two kinds, namely one-way encryption algorithm and two-way encryption algorithm, the latter is the most commonly used , it will be able to encrypt understand directly expressly data for non-understand directly ciphertext data, in time of need, and can encrypt these encryption later ciphertext data declassified for original expressly data by using certain algorithm. Two-way encryption is suitable for secret communications. And one-way encryption opposite, it can only encrypt expressly data, and not encrypt ciphertext data for the original expressly data. This is of great significance, for instance, to encrypt system user information (such as user password) at database of the software system. When the user registered a new account, the user password information is not directly saved to the database, but after one-way encrypted save again, so that, even if the account information was leaked, others also cannot get the corresponding user password, of course also constitute no account theft of purpose.

Md5 encryption algorithm is one of one-way encryption algorithm. It has two important characteristics, the first is any two paragraphs expressly data, after encrypted the data is not the same plaintext; The second is any one data, encrypted later, the ciphertext is always the same data.

2 MD5

Md5's full name is the Message-Digest Algorithm 5 (information-the Algorithm), in the 90 s developed by the Mit Laboratory For Computer Science and Ronald Rivest of Rsa Data Security Inc's, The preliminary version includes the md2、md3 and md4. Its role is to let the large capacity information be compressed into a confidential format before sign the private key using digital signature software.

MD5 handle input information by 512 bits group, and each group is divided into sixteen 32-bit seats group, after a series of processing, the output of the algorithm is grouped by four 32-bit child group, that will create a 128 a hash value after cascade.

In the MD5 algorithm, the first, need to filling the information, the result of remainder making its bytes length of 512 is equal to 448. Therefore, Bytes Length (Bits Length) of the information will be extended to $N * 512 + 448$ Bits, namely $N * 64 + 56$ Bytes, N for a positive integer. Filling the method is as follows, the back of the information is filled with a 1 and countless 0, until they meet the condition of above. And then in back of the results, appending the length of information before filling said by 64 bit binary. After this two-step process, now bytes length of the information is $N * 512 + 448 + 64$, that is $(N + 1) * 512$, namely the length of 512 is just an integer multiple. The reason for doing this is to meet the request of information length in the treatment behind. MD5 there are integers parameters of four 32-bit is called link variables (Chaining Variable) , they are: A = 0 x01234567, B = 0 x89abcdef, C = 0 xfedcba98, D = 0 x76543210. after setting up the four links variables, began to enter the fourth cycle computing algorithm, the number of circulation is the number of 512 bit information group. Will copy above four links variables to other four variables: A to a, B to b, C to c, D to d. The main loop have four-wheel (MD4, only three rounds), each wheel cycle are very similar. For the first round, it has 16 times operation. Every operation is a nonlinear function operations on the three of a, b, c and d , then add the fourth variables to the results. then move the results an unspecified number to the right ring, and add. Finally one of a, b and c or d is replaced by the result. The following is the four nonlinear function used in each operation (each wheel a).

$$F(X, Y, Z) = (X \wedge Y) \vee (\neg X \wedge Z) \quad (1)$$

$$G(X, Y, Z) = (X \wedge Z) \vee (Y \wedge \neg Z) \quad (2)$$

$$H(X, Y, Z) = X \oplus Y \oplus Z \quad (3)$$

$$I(X, Y, Z) = Y \oplus (X \vee \neg Z) \quad (4)$$

Among them, \oplus , \wedge , \vee , \neg respectively is XOR, AND, OR, NOT symbols.

If corresponding position of X, Y and Z is independent and uniform, so each of the results also should be independent and uniform. F is a operation function by bit . That is, if X, then Y, or Z. Function H is parity operators by bit. All this is done, A, B, C, D will be add respectively a, b, c, d. Then use the next packet data to continue working

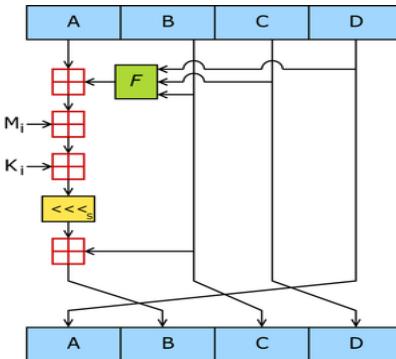


Fig. 1. A MD5 operations

algorithm, the output of the last is cascade of A, B, C and D . The final A, B, C, D is output results, A is low, D for high, DCBA composite 128-bit output. Md5 operation diagram, such as Figure 1.

By similar 64 times a cycle, divided into four groups 16 times. F is a nonlinear function; one function operate once. M_i said a input data of 32-bit bytes, K_i says a 32-bit constant, used to accomplish every different calculation.

3 Security and Defect Analysis

MD5 is a secure hash algorithm, the input two different expressly won't get the same output value, according to the output value, can't get the original expressly, namely the irreversible process; So no ready algorithm to decrypt MD5,only use exhaustively method, form a one-to-one mapping table about the possible expressly, with their hash value after MD5 hashing algorithm, find out the corresponding original plaintext of cracked the password through the match from mapping table.

The information systems or the website system, MD5 is mainly used to encrypt password of user registration, for ordinary strength password encryption, can be cracked through the following three ways:

(1) online inquires the password. Some online MD5 value inquires site provide the server, input MD5 password value, if exist in the database, then can quickly get the password value.

(2) use MD5 cracked tools. The network has many special software to MD5 crack, by setting up the dictionary to crack.

(3) obtain or resetting user's password through the social engineering.

So simple MD5 encryption is not a way to get to the absolute safety, because the average MD5 encryption has a variety of brute force way, so if you want to ensure that information system or site security, the need to reform MD5, strengthen its security, this paper is improving on the basis of the MD5 encryption algorithm.

4 The Improved Encryption Method Based on MD5 Algorithm

This method mentioned in the article: on the basis of encrypting expressly data by using md5 encryption algorithm, change the plaintext, then intercepte part data and discard, and then use a random function to fill discarded data, and the whole process does not change digits after the md5 encryption. The encryption process using algorithm are described below:

(1) for encrypting expressly data---password by md5, get ciphertext---getMD5ofStr (password).

(2) intercept encryption ciphertext using intercepted function. intercept number position beginning from the beginnumber position, get password A , including A =getMD5ofStr (password). substring (0, beginnumber).

(3) get B which is the string of intercepting ciphertext in (1) from the beginnumber + number position, namely B = getMD5ofStr (password).substring (beginnumber + number).

(4) use random function---getRandomChar (int number) to fill the value intercepted number.

(5) the transformed password value: encrypt_password = A + getRandomChar (int number) + B.

Variable description:

password: client submit original password

encrypt_password: the password after process

beginnumber: began to intercept the position of the original password(0 ~ password length-1)

number: discarded data length or the length of the string generate by random function.

Decryption process is similar with encryption process, first to encrypt the input data, then from beginnumber place intercepted first half, get A ', half of the, get B ', and read part A and part B of the password from the database , finally if A = A 'and B = B', and is considered the password of user input is matching with password in the database.

key code of improvement methods mentioned in this article its as follows:

```
public String getRandomChar(int size)
{
    String s="";
    char[] codeSequence = { 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'T', 'J', 'K', 'L', 'M',
'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z', '0', '1', '2', '3', '4', '5', '6', '7', '8', '9' };
    int m_intRandlen = size;
    for (int i = 1;i<=m_intRandlen;i++)
    {
        Random ran=new Random();
        char c=codeSequence[ran.nextInt(36)];
        s=s+c;
    }
}
```

```

    return s;
}
public String encrypt_password(int beginnumber,int number,String password)
{
    String ss=this.getMD5ofStr(password);
    String sleft=ss.substring(0,beginnumber);
    String sright=ss.substring(beginnumber+number);
    String smiddle=getRandomChar(number);
    String newpassword = sleft+smiddle+sright;
    return newpassword;
}

```

5 Application of Improved Md5 Algorithm

This paper proposes online shopping system based on the multi-storey structure Ajax technology. The system is divided into several modules, use the above framework for the design of loosely coupled, and can allow multiple data sources, interactive EIS, and to achieve client and server-side data asynchronous interaction, improving the maintain ability and scalability of the J2EE overall frame-work.

The system design framework diagram is shown in Fig 2.

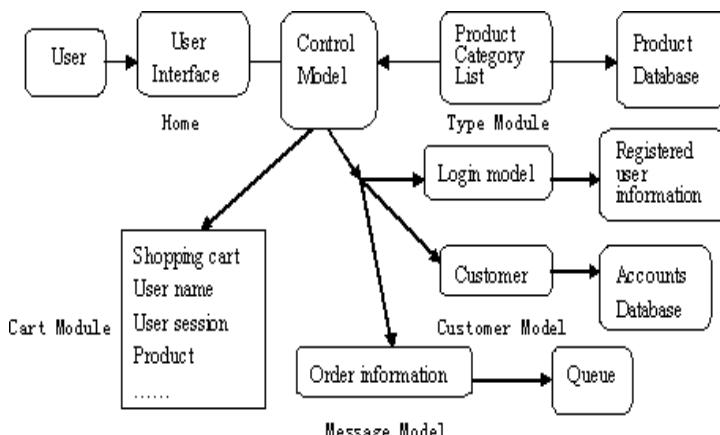


Fig. 2. System Framework

The major modules and function as follows: 1) Control Model: used for the circulation of requests to the various business processing logic, Jump screen control, as well as dealing with the corresponding components of the Login and user registration. the module is realized and expanded by WAF (Web Application Framework). 2) Cart Module: the tracking main shopping process of users. 3)Login Model:the need for users to log some pages.4) Message Model: asynchronous transfer

orders from businessmen to center. 5) Type Module: In accordance with user requirements for a category of view. 6) Customer Model: customer information, address, credit card, the associated pattern. Among them, the control module controls all the interactive and implementation, each user session references to a Cart. JSP achieve User interface; Structure of control module, Cart module, landing module, message module etc use JavaBean, Servlet and Struts in Action, and Ajax technology is introduced to Struts, which can create dynamic Web pages; the role of Hibernate is that the database will be converted to the object persistence layer in the category, login module, to achieve the mapping between objects and relations, in order to Application procedures call.

In this paper, with user registration process as an example. According to the requirements , users fill the information as shown in Fig 3, and submit it, then the submitted information is handled by the corresponding servlet class, such as password which is ljr123\$%, is encrypted with improved encryption method above mentioned, to be stored database table---users, results as shown in Fig 4. From it, we cannot know what is original data of the user's identity card number and password, Thus, effective guarantee of private information security.

The screenshot shows a web browser window titled "MyEclipse Web Browser" displaying a JSP page at "http://localhost:8080/ss/user_reg.jsp". The page has a green header "New User Registration" and a red sub-header "Welcome you,let us join hands !". It contains several input fields for user registration, including "User Name" (ljr), "Password" (*****), "Confirm_Password" (*****), "Real Name" (empty), "Sex" (empty), "E-Mail" (ljrokies@163.com), "Telephone" (empty), and "Verification code" (empty). Below the form is a CAPTCHA field with the value "1 I n N" and a link "See? for one". At the bottom are "submit" and "reset" buttons, and a red warning message "Be careful: * must be written!".

Fig. 3. New User Registration

The screenshot shows the MySQL Workbench interface with a table named "Users". The table has columns: id, username, password, cpassword, realname, sex, email, and tel. A single row is selected, showing the following data: id=1, username=ljr, password=B70BN85VNE2B9FD3FB498C1F7FC43142, cpassword=B70BN85VNE2B9FD3FB498C1F7FC43142, realname=, sex=m, email=ljrokies@163.com, and tel=.

Fig. 4. Users table

6 Summary

Someone had also puts forward to modify the function or the variable in the md5 encryption algorithm, in order to strengthen the security of the original md5 algorithm used, but this method changes the md5 antiderivative or variables, unable to verify in intensity the changed md5 algorithm is consistent with the original algorithm. The method in this paper is, on the basis of the original md5 encryption, intercept certain digit string to ciphertext, and fill it by random number, though the final ciphertext is after md5 encryption, but its value is not, so through the md5 conventional crack method is never cracked the primitive code value, so as to ensure the safety of the data. Although there are many attack methods, such as SQL injection, cross site attack, and can easily get value in the database, through this method encryption, before the leak of websites or system code, the data is relatively safe, so there are some reference bonus.

References

1. Zhang, H.-H., Qi, W.-Y., et al.: Design and implementation of MD5 algorithm based on java. Journal of ShenYang Normal University (Natural Science) 27(1), 26–29 (2009)
2. Chen, S.-H., Zhai, X.-N., Yan, N., Su, G.-X.: Cryptanalysis of deciphering MD5 algorithm. Computer Engineer and Application, 99–102 (2010)
3. Weng, J., Wu, Q., Yang, C.-Q.: MD5 crack algorithm Based on the OpenCL. Computer Engineering, 65–67 (2011)
4. Sun, W.-G., Li, H.-R.: MD5 Algorithm in Data Security and Safety of the Application. Microcomputer Application, 158–160 (2010)
5. Zhang, S.-L., Xing, G.-B., Yang, Y.-X.: Improvement and Security analysis on MD5. Computer Applications, 948–949 (2009)
6. Le, D.-G., Chang, J.-Y.: Implementation of MD5 Fast Decryption Algorithm Based on Graphic Processing Unit. Computer Engineering, 34–36 (2010)
7. Zhang, X.-H., Zhen, T., Bao, H.: Security Registration System Based on the DES algorithm and MD5 algorithm. Computer Digital Engineering, 132–133 (2010)
8. Zhang, Y.-Z.: Md5 Algorithm Research. Computer Science, 207–209 (2008)
9. Yue, J.-B.: Application of Dynamic two-factor authentication technology based on MD5 algorithm in enterprise ERP system. Automation and Instrumentation, 51–52 (2010)
10. Peng, W.-B.: Theory and Applicationthe of MD5 algorithm. Information Network, 44–46 (2005)

Adaptivity in Location-Based Services

Rui Zhou, Wensheng Guo, and Nan Sang

School of Computer Science and Engineering,
University of Electronic Science and Technology of China

Abstract. Location-based services have gained wide employment in the recent years. In addition to user location, the services can be adapted to the context such as user interests, preferences, personal profiles, mobile devices and time. Even the location system per se can be adaptive too, which incorporates multiple available positioning methods and returns an fused and improved location estimation. This paper first presents the concept of adaptive location system and discusses its implementation with sensor fusion algorithms. Then the paper focuses on how to apply context-awareness and adaptivity to location-based services. At the end of the paper the implementation of adaptivity in a real location-based application is presented.

Keywords: location-based service, adaptivity, context.

1 Introduction

Computing and networking technologies are changing the way that people learn, work and live. Positioning technology makes it possible to keep track of people and mobile devices and thus gives rise to location-based services, which are accessible with mobile devices and make use of the geographic locations of the mobile devices to tailor the services [14]. Location-based services integrate the virtual cyberspace with the real physical world and provide only the information that the user really needs, thus spare the time and effort to find useful information from the vast sources.

Location-based services can be applied in many aspects of our daily lives, and they do have gained much attention and employment over the last years. Navigation is one of the most important applications. With knowledge of a user's real time location, a navigation system can show his or her exact location on the map and give turn-by-turn direction to his or her desirable destination, even able to take into account the real time traffic situation and the transportation connections. Upon receiving a mobile user's request, mobile yellow pages are able to provide the user with the information in the vicinity. Modern tourism systems can present the information about the nearby attractions or notify the user when he or she is approaching a certain attraction. By tracking, friends can find each other, parents can monitor the exact whereabouts of their young children, corporations can locate their personnel within a certain range, logistic companies are able to know where their products are at any time. E911 in

US (and E112 in EU) ensures the emergency assistance to automatically locate the citizens dialling 911 (or 112) and arrive at the spot promptly. Surveilling patients in hospitals allows the doctors and nurses to monitor the conditions of the patients and provide prompt medical assistance. There are much more applications of location-based services, which is out of the scope of the paper.

In addition to being tailored to user locations, location-based services are often adaptive to other context, such as user interests, user preferences, mobile device used, time, date, weather, etc. Treating each user as an individual and provides only the necessary information helps to improve the efficiency and spares the users from repetitive work. In one word, being adaptive is to provide the appropriate service in the appropriate way at the appropriate place and the appropriate time. As we are in an information explosion age, adaptivity is of vital importance to the success of location-based services. Adaptivity has wide application in traditional desktop educational systems [3,2,10,9], and is spreading to location-based services [5,13,20,11,1]. We implemented an adaptive location-based guiding system in botanical gardens. In this paper we elaborate on the adaptivity techniques that are adopted and the implementation in the real application.

2 Adaptivity

2.1 Adaptive Location System

Over the years many systems have tackled the problem of positioning, but none of them takes perfect measurements or works well in all situations. GPS is currently the predominant approach to outdoor navigation, which makes use of the constellation of 24 satellites, controlled by the US Department of Defense (DoD), and time of arrival of the signals from the satellites to determine the positions of GPS receivers. Its average accuracy is about 5-10 meters. However, GPS has difficulties in providing accurate position under poor signal conditions, e.g. in buildings, urban canyons, and RF-shadowed environments. Wi-Fi positioning is accepted as a common indoor positioning approach, which makes use of the signals of 802.11 wireless networks to determine user locations. It can achieve an average accuracy of 2-5 meters indoors [17] and 10-40 meters outdoors [4]. In the cellular networks, positions of mobile phones can be derived from the signal measurements using time of arrival (TOA), time difference of arrival (TDOA) or angel of arrival (AOA). But the positioning accuracy is about 300-500 meters, which is quite coarse. Other positioning approaches include RFID, UWB, infrared, ultrasound and inertial navigation system.

Each positioning approach has its pros and cons. No one suits well in all situations. On the other hand, users often carry different mobile devices, equipped with different positioning gadgets. An adaptive location system is to adaptively select positioning approaches according to the available postioning gadgets and combine them to provide an improved location service using sensor fusion algorithms [7,8]. Bayesian filtering [6] is a common solution implementing sensor fusion in the robotics community. It can probabilistically estimate the state of

a dynamic system from noisy sensor observations. For location estimation, the state is the location of a mobile device, and the observations are measurements from the multiple positioning devices. Suppose z_1, z_2, \dots, z_t is a sequence of time indexed measurements, and x_t represents the user location at time t , the key idea of Bayesian filtering is to estimate the probability distribution over x_t , called *belief*, conditioned on all measurements [6]:

$$Bel(x_t) = p(x_t | z_1, z_2, \dots, z_t)$$

It answers the question: “For all possible locations, what is the probability that the mobile device is at location x_t , if the history of measurements is z_1, z_2, \dots, z_t ? ”.

The complexity of computing $Bel(x_t)$ grows exponentially over time as the number of the measurements grows. In order for efficient computation, Bayesian filters assume the dynamic system is a Markov process, which implies [6]: the observations only depend on the current state of the system; the state x_t only depends on the state x_{t-1} . Under the assumptions, the recursive state estimation by Bayesian filters can be described as prediction and correction phases [6]:

- Prediction: When a new measurement is made, the location is predicted as:

$$Bel^-(x_t) \leftarrow \int p(x_t | x_{t-1}) Bel(x_{t-1}) dx_{t-1}$$

- Correction: The predicted location is corrected using the new measurement:

$$Bel(x_t) \leftarrow \alpha_t p(z_t | x_t) Bel^-(x_t)$$

Here, $Bel^-(x_t)$ is the predictive belief. $p(x_t | x_{t-1})$ is called motion model which describes how the location of the mobile device changes over time. $p(z_t | x_t)$, called the likelihood model, is the likelihood of taking measurement z_t if the location of the mobile device is x_t . Each positioning technology needs a likelihood model characterizing it. α_t is a normalizing constant which ensures the distribution $Bel(x_t)$ sums to 1. To implement Bayesian filters, the belief $Bel(x_t)$, the motion model $p(x_t | x_{t-1})$ and the likelihood model $p(z_t | x_t)$ must be specified.

2.2 Adaptation to Location

For a location-based service, the content or service must be adapted to the user location, which is usually achieved via location-dependent content queries. There are a few common types of queries depending on the application scenarios and the users’ requests. A planar point query [15] returns the geographic objects that contain, surround or refer to a specific location. Using the user’s location as the querying point, it can be used to query the region that the user is currently in and its relevant information. Here are some applications of planar point queries. Many positioning techniques return the geographic coordinates as the locations, while many applications require the geographic symbolic locations, e.g. city,

zone and street. The conversion from a geographic coordinate to a symbolic location often involves planar point queries. Information such as local traffic, local weather forecast, attractions and public facilities is valid within a certain region. The region can be derived from planar point queries using the user's location as the querying point. A window query [16] returns the geographic objects or points of interest that intersect or are contained in the specific region. It can be used to return the points of interest that are within a certain distance to a user, when using the user's location and the distance to specify the querying window. A nearest neighbours query [16] returns the geographic objects or points of interest that are closest to a specific location, that is, the nearest points of interest to the user when using the user's location as the querying point. A path query asks for the route from a specific location to another specific location. For example, what is the shortest route from A to B? or what is the fastest route from A to B considering real time traffic situation?

2.3 Adaptation to User Interests

Interests are the most commonly used characteristic for adaptive location-based services. Knowing the user interests, the service provider can provide only the information that the user is interested in. The more interested the user is, the more detailed information he or she will receive. Modelling of the user interests is via an overlay on the domain model. The domain model has a network structure, in which the nodes are the points of interest (or concepts) in the domain and the edges between the nodes are their relationships. Suppose

$$P = \{p_i | i = 0, 1, \dots, n\}$$

represents the set of the points of interest in the domain, p_i stands for the i -th point of interest, and

$$E = \{e_{ij} | i = 0, 1, \dots, n; j = 0, 1, \dots, n\}$$

represents the set of the relationships between the points of interest, e_{ij} stands for the relationship between the i -th point of interest and the j -th point of interest, then the domain model can be expressed as

$$D = \{P, E\}$$

The user interest model can be expressed as

$$M = \{(p_i, v_i) | i = 0, 1, \dots, m; m \leq n; p_i \in P\}$$

p_i is the i -th point of interest, v_i is the corresponding interest value, which can be qualitative values (e.g. low-intermediate-high), or numeric values (e.g. 0–10).

The user interest model is empty initially, and gets updated continuously during the course of interaction. The interest level of a point of interest will get increased if the user chooses to visit it. The longer the user stays with the point

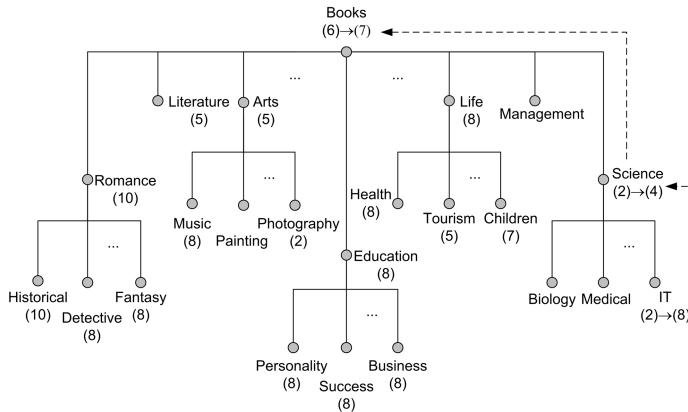


Fig. 1. An example of user interest model in the domain of books

of interest, the higher the interest level will be. If the user is interested in a point of interest, he or she may be interested in the related points of interest, the process is called interest propagation. When updating the interest level of a point of interest, it is often necessary to update the propagated points of interest. Interest propagation is a complicated process, which needs the assistance of the domain experts and evaluation and adjustment during operational use.

Figure 1 is an example of user interest model in the domain of books. The nodes are the classifications of books, the edges connecting the nodes are the relationships, which is *kind-of* in this scenario. The numbers beside the nodes are the interest levels, which is a number between 0 meaning no interest at all and 10 meaning extremely interested. If the interest level of a node is unknown, there is no number beside it. Thus, the user interest model of Fig. 1 can be represented as:

$$M = \{(Books, 6), (Romance, 10), (Literature, 5), (Arts, 5), \\ (Education, 8), (Life, 8), (Science, 2), (Historical, 10), (Detective, 8), \\ (Fantasy, 8), (Music, 8), (Photography, 2), (Personality, 8), (Success, 8), \\ (Business, 8), (Health, 8), (Tourism, 5), (Children, 7), (IT, 2)\}$$

When the user interest level for *IT books* is increased from 2 to 8, the interest level for *Science books* gets increased from 2 to 4, and the interest level for *Books* gets increased from 6 to 7.

2.4 Adaptation to Other Context

The context is more than just location and user interests. Adapting the service to users' personal profile, such as name, gender, title, nationality, vocation, and things alike, will make the service more personal and thereby enhance users' experience. Preferences specify how a user likes the content to be presented,

which may be the multimedia types that the user prefers, the notification mode when new information is available (e.g. automatically start or manually control), the presentation mode (e.g. all-in-one page, multiple pages, or slideshow), and so forth. Adaptation should take user preferences into consideration. Different mobile devices, such as laptops, pads, or mobile phones, have different operating systems, screen sizes, software and capabilities, which makes the users' experiences quite different. It is necessary to adapt the content to the mobile device. As there are many different features of mobile devices and the number of the combinations will be huge, it is impractical to realize content adaptation for each combination. The common solution is mapping the combinations of features of the mobile devices to a few stereotypes and adapting the service to the stereotype. Location-based services should also adapt to the context elements like time, date, season, weather, lighting and noise. For example, in different seasons or under different weather conditions, different content will be selected; if it is too light to see the screen clearly, audio will be used to present information; if it is too noisy, images and texts will be more appropriate [18].

3 Implementation

We have implemented an adaptive multimedia guiding system to be used in botanical gardens covered with an 802.11 wireless network [19]. The system is able to keep track of users equipped with GPS receivers and Wi-Fi adapters and provide adaptive multimedia content of the nearby plants to them. The system consists of a botanical web server offering the adaptive botanical guiding service, a location server hosting the adaptive location system which fuses GPS and Wi-Fi fingerprinting, a content server hosting the content database, and web browsers as the clients. To support location-awareness, the web browsers are enhanced with a specific extension to retrieve location-related parameters from Wi-Fi adapters and GPS receivers. User interests and user knowledge about botany are modelled via overlay models as a set of (*botanical name, interest level*) pairs and a set of (*botanical name, knowledge level*) pairs. Other context

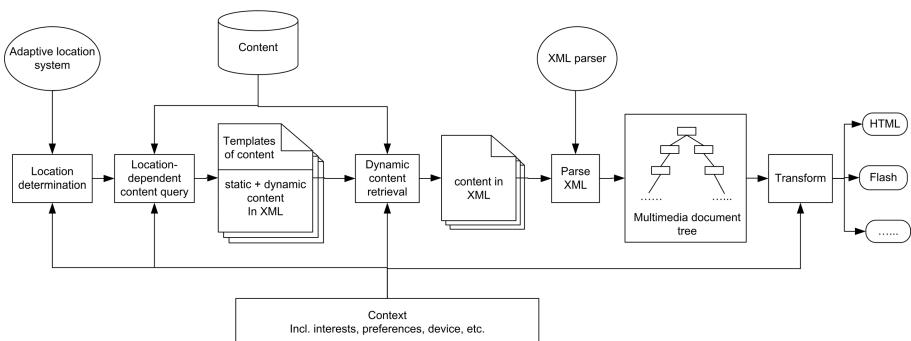


Fig. 2. Implementation of adaptivity

elements such as preferences, personal profile and mobile device are modelled as *(key, value)* pairs.

Implementation of adaptivity in the system is illustrated in Fig. 2. The user location is first determined by the adaptive location system and then a location-dependent query is performed to find out the services and/or information in the vicinity of the user. Based on the predefined templates, which consists of static part and dynamic part, and the context, the system chooses the appropriate content and composes them to a multimedia document tree. The multimedia document tree is an abstract model that contains the media elements to be presented and describes the spatial and temporal relationships between them and how they will be presented. It can be transformed to different formats such as HTML or Flash, which makes possible dynamic authoring of adaptive multimedia presentations in different formats and for different devices, hence provides more flexibility [12].

4 Conclusions

Adaptivity is a natural feature of location-based services. The services should be tailored not only to user locations, but also to user interests, preferences, mobile devices, and etc. In addition to service adaptivity, the location system can also be adaptive, which incorporates different positioning methods and selects the most suitable ones in a specific situation to offer a more flexible and robust location service. Our concrete implementation of a real application demonstrates the feasibility of the presented methods and how to implement adaptivity in a location-based service.

References

1. Ahn, C., Nah, Y.: Design of location-based web service framework for context-aware applications in ubiquitous environments. In: 2010 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC), pp. 426–433 (June 2010)
2. de Bra, P., Aerts, A., Berden, B., de Lange, B., Rousseau, B., Santic, T., Smits, D., Stash, N.: AHA! the adaptive hypermedia architecture. In: Proceedings of the 14th ACM Conference on Hypertext and Hypermedia (Hypertext 2003), Nottingham, UK, pp. 81–84 (August 2003)
3. Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): The Adaptive Web—Methods and Strategies of Web Personalization. Springer, Heidelberg (2007)
4. Cheng, Y.C., Chawathe, Y., LaMarca, A., Krumm, J.: Accuracy characterization for metropolitan-scale Wi-Fi localization. In: Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys 2005), Seattle, WA, USA, pp. 233–245 (June 2005)
5. Cheverst, K., Mitchell, K., Davies, N.: The role of adaptive hypermedia in a context-aware tourist guide. Communications of the ACM 45(5), 47–51 (2002)
6. Fox, D., Hightower, J., Liao, L., Schulz, D., Borriello, G.: Bayesian filtering for location estimation. IEEE Pervasive Computing 2(3), 24–33 (2003)

7. Gwon, Y., Jain, R., Kawahara, T.: Robust indoor location estimation of stationary and mobile users. In: Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (Infocom 2004), Hong Kong, China, vol. 2, pp. 1032–1043 (March 2004)
8. Hightower, J., Borriello, G.: Particle Filters for Location Estimation in Ubiquitous Computing: A Case Study. In: Davies, N., Mynatt, E.D., Siio, I. (eds.) UbiComp 2004. LNCS, vol. 3205, pp. 88–106. Springer, Heidelberg (2004)
9. Kobsa, A., Koenemann, J., Pohl, W.: Personalised hypermedia presentation techniques for improving online customer relationships. *The Knowledge Engineering Review* 16(2), 111–155 (2001)
10. Papanikolaou, K.A., Grigoriadou, M., Kornilakis, H., Magoulas, G.D.: Personalizing the interaction in a Web-based educational hypermedia system: The case of INSPIRE. *User Modeling and User-Adapted Interaction* 13(3), 213–267 (2003)
11. Richter, K.F., Dara-Abrams, D., Raubal, M.: Navigating and learning with location based services: A user-centric design. In: Proceedings of the 7th International Symposium on LBS and Telecartography, pp. 261–276 (2010)
12. Scherp, A.: A Component Framework for Personalized Multimedia Applications. Ph.D. thesis, Department of Computing Science, School of Computing Science, Business Administration, Economics and Law, Carl von Ossietzky University of Oldenburg, oIWIR, Oldenburg, Germany (February 2007)
13. Schmidt-Belz, B., Nick, A., Poslad, S., Zipf, A.: Personalized and location-based mobile tourism services. In: Workshop on Mobile Tourism Support Systems in Conjunction with Mobile HCI 2002, Pisa, Italy (September 2002)
14. Virrantaus, K., Markkula, J., Garmash, A., Terziyan, V.Y., Veijalainen, J., Kata-sonov, A., Tirri, H.: Developing GIS-supported location-based services. In: Proceedings of the 2nd International Conference on Web Information Systems Engineering (WISE 2001), Kyoto, Japan, vol. 2, pp. 66–75 (December 2001)
15. Xu, J., Zheng, B., Lee, W.C., Lee, D.L.: The D-tree: An index structure for planar point queries in location-based wireless services. *IEEE Transactions on Knowledge & Data Engineering* 16(12), 1526–1542 (2004)
16. Zhang, J., Zhu, M., Papadias, D., Tao, Y., Lee, D.L.: Location-based spatial queries. In: Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2003) (2003)
17. Zhou, R.: Enhanced wireless indoor tracking system in multi-floor buildings with location prediction. In: Proceedings of the 12th International Conference of European University Information Systems (EUNIS 2006), Tartu, Estonia, pp. 448–453 (June 2006)
18. Zhou, R., Rechert, K.: Personalization for location-based e-learning. In: Proceedings of the 2nd IEEE Conference on Next Generation Mobile Applications, Services, and Technologies (NGMAST 2008), Cardiff, Wales, UK, pp. 247–253 (September 2008)
19. Zhou, R., Schneider, G.: Location-Based Botany Guide: A prototype of Web-based tracking and guiding. In: Proceedings of the 2nd International Conference on Mobile Wireless Middleware, Operating Systems, and Applications (Mobilware 2009), Berlin, Germany, pp. 72–86 (April 2009)
20. Zhu, T., Wang, C., Jia, G., Huang, J.: Toward context-aware location based services. In: 2010 International Conference on Electronics and Information Engineering (ICEIE), pp. 409–413 (August 2010)

An Evolution Model of Emotional Internet Public Opinion with Informed Marks

Wu Yu¹, Hao Wang¹, Xian Wang¹, and Jianbo Li²

¹ Institute of Web Intelligence, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China

² Chongqing Educational Examinations Authority, Chongqing 401147, China
wuyu@cqupt.edu.cn

Abstract. The existing models of Internet public opinion don't take into account the effects of individual's informed status and emotional tendency on the formation of public opinion. In response to this problem, based on emergent computing theory, an evolution model of emotional Internet public opinion with informed marks is proposed. In this model the cyber citizen are abstracted as agents and some parameters such as emotion tendency and initial informed marks are added. In order to accord with the propagation's law of public opinion, the interaction rules adopted by current models are improved. Compared with the existing model, this model analyzes the newly added parameters' influence on the public opinion.

Keywords: the evolution of public opinion, emergence, emotional Internet public opinion, small-world network.

1 Introduction

Consensus referees to the set of social public opinion and its propagation forms are diverse. Recently, mass media is playing an increasingly important role in spreading public opinion, which profoundly affects politics, economy, technology and culture. Especially, micro-blog attracts more and more attention as a new transmission platform [1]. Therefore, studying the propagation laws of public opinion in media has important theoretical value and practical significance.

Research on evolution of public opinion has been a hot spot [2-6]. Some famous models include Sznajd model [7], Deffuant model [8] and Krause-Hegselmann model. In these models, the individual's opinion is based on its neighbours' views. In this way, the "disorder-order" evolution of public opinion is imitated. On the other hand, public opinion is a kind of dynamical collective behaviour formed by numerous individuals. From this perspective, public opinion belongs to the area of emergent computation, which is an idea exhibited by multi-agent systems, and describes the process of enormous simple agents forming complex behaviors by cooperating. However, at present few scholars research the public opinion from this point of view, or the hypotheses are too idealistic and lack practical applicability. For instance, with

ideas of emergent computing, Wu et.al studied the evolution process of public opinion, but didn't consider the factors such as individuals' emotion, so their work couldn't repeat the evolution process of public opinion in real situations.

Besides, in virtual environment, the individual's psychology and emotion also have an effect on his cognition and thinking, eventually influencing the final trend of public opinion. Therefore, Peng proposed the concept of "emotional opinion in Internet" and analyzed its causes and guiding strategy. Later, some scholars did some researches on emotional Internet public opinion, but mostly from the perspective of social science. Also, there are some researchers who think about the emotional tendencies in public opinion, but the influence of emotion on the public opinion is not analyzed quantitatively.

Based on above background, combined with the idea of emergent computing, an evolution model of emotional Internet public opinion with informed marks is proposed (IS&NPEO), in which the informed status of each individual and his emotions are both considered. Then, the added factors' impacts on the evolution are analyzed in experiments.

2 Previous Work

With the idea of emergent computation, Wu et.al established the model of net public opinion based on small-world (NPO-NW) model, in which every cyber citizen is abstracted as an agent. The system is defined as $S = \langle A, P, f \rangle$. Details are given in the following [11]:

$A = (A_1, A_2, \dots, A_i, \dots, A_N)$, where A_i represents the i agent, $i = 1, 2, \dots, N$, and N is the number of agents;

$A_i = \{(op_i, pe_i) | 0 < op_i < 1, 0 < pe_i < 0.5\}$, where op_i is the opinion value of A_i , while pe_i is the uncertainty value to its opinion;

P is the collection of control parameters in this system and $P = \{(d_{\min}, d_{\max}, tr) | 0 < d_{\min} < d_{\max} < 1, 0 < tr < 1\}$, where tr is the overall trust level between agents, while d_{\min} and d_{\max} are the opinion threshold, representing the lower and upper levels of the opinion's difference between A_i and A_j respectively;

f is the interaction function between agent i and j . When $d_{ij} < d_{\min}$ or $d_{ij} > d_{\max}$, $f(i, j) = 0$, which means that when differences in opinion between A_i and A_j are very small or very large, they don't exchange their views. Beyond that, A_i and A_j shall exchange their views via:

$$f(i, j) = \begin{cases} +tr \times |op_i \times pe_i - op_j \times pe_j|, & d_{ij} / pe_j < 1 \\ -tr \times |op_i \times pe_i - op_j \times pe_j|, & d_{ij} / pe_j \geq 1 \end{cases} \quad (1)$$

Small-world network is used to represent interpersonal relationship, in which each node represents an agent. Each agent's status is updated synchronously. Finally, the opinion and persist of A_i and A_j at time $t+1$ are calculated as:

$$\begin{cases} op_i(t+1) = op_i(t) + f(i, j) \times (|op_i(t) - op_j(t)|) \\ pe_i(t+1) = pe_i(t) + f(i, j) \times (|pe_i(t) - pe_j(t)|) \end{cases} \quad (2)$$

With various combinations of tr , d_{\max} and d_{\min} , the results of simulations demonstrate that many trends of public are formed, such as single-opinion, double-opinion, multi-opinion, which is of great importance to the research on public opinion.

3 The Evolution Model of Emotional Internet Public Opinion with Informed Marks

Based on NPO-NW model, adding the informed marks and emotion tendency, an evolution model of emotional Internet public opinion with informed marks (IS&NPEO) is constructed.

3.1 The Description of IS&NPEO Model

IS&NPEO model is a system composed of N agents, defined as $S = \langle A, P, F \rangle$, details of which are given as follows:

$A = (A_1, A_2, \dots, A_i, \dots, A_N)$, where A_i represents agent $i, i = 1, 2, 3, \dots, N$, and N is the number of agents;

$A_i = \{(op_i, e_i, in_i) | 0 \leq op_i \leq 1, 0 \leq e_i \leq 1, in_i = 0, 1, i = 1, 2, 3, \dots, N\}$, where op_i is the opinion value of A_i , e_i is the emotion tendency value of A_i , and in_i is the informed status of A_i ;

P is the set of control parameters in this system and $P = \{d_{\min}, d_{\max}, tr | 0 < d_{\min} < d_{\max} < 1, 0 < tr < 1\}$, where tr is the overall trust level between agents, while d_{\min} and d_{\max} are the opinion threshold, representing the lower level and upper level of the difference of opinion between A_i and A_j respectively;

The difference of opinion between A_i and A_j is defined as $d_{ij} = |op_i - op_j|$

$F = [f_{ij}]_{N \times N}$ is the matrix of emotional functions, where f_{ij} is the emotional function and defined as:

$$f_{ij} = f(i, j) = \begin{cases} 0 & , d_{ij} < d_{\min} \text{ or } d_{ij} > d_{\max} \\ +I_{ij} \times tr \times |op_i \times e_i - op_j \times e_j| & , d_{\min} \leq d \leq d_{\max} \text{ and } e_i > d_{ij} \\ -I_{ij} \times tr \times |op_i \times e_i - op_j \times e_j| & , d_{\min} \leq d \leq d_{\max} \text{ and } e_i \leq d_{ij} \end{cases} \quad (3)$$

where $I_{ij} = \frac{in_j}{in_i + in_j + \varepsilon}$ is the informed impact-factor and ε is a constant tending to zero, called informed correction constant;

The interpersonal network is constructed by using the small-world model. All the agents update their status synchronously in each loop, according to the following formula:

$$\begin{cases} op_i(t+1) = op_i(t) + f(i, j) \times (|op_i(t) - op_j(t)|) \\ e_i(t+1) = e_i(t) + f(i, j) \times (|e_i(t) - e_j(t)|) \\ in_i(t+1) = 1, in_i(t) = 1 \text{ or } in_j(t) = 1 \end{cases} \quad (4)$$

In contrast with NPO-NW model, the agent's informed status is taken into account as well as the influence of emotion on the agent's opinion. The uncertain factor pe ($pe \in [0, 0.5]$) in NPO-NW model is amended to emotional tendency e ($e \in [0, 1]$). So the range of changes of the agent's opinion is controlled, making up for the deficiencies of NOP-NW model in the symbol of f_{ij} . Moreover, when A_i and A_j are both informed, that is $in_i = 1$ and $in_j = 1$, the emotional function of the proposed model is

$$f(i, j) = \begin{cases} +0.5 \times tr \times (op_i \times e_i - op_j \times e_j) & , e_j > d_{ij} \\ -0.5 \times tr \times (op_i \times e_i - op_j \times e_j) & , e_j \leq d_{ij} \end{cases} \quad (5)$$

Therefore, it can be seen that NPO-NW model is a special case of the proposed model. And the proposed model is an extension of the NPO-NW model.

3.2 Algorithm Description

Several ratios are used in the algorithm, including the initial network informed rate p ($0 \leq p \leq 1$), the rates of the cybergitizens' emotion inclination q_1 , q_2 ($0 \leq q_1 \leq 1$, $0 \leq q_2 \leq 1$), and the rate of opposing arguments k ($0 \leq k \leq 1$), respectively defined as:

$$p = \frac{s_0}{s}, q_1 = \frac{s_n}{s_0}, q_2 = \frac{s_p}{s_0}, k = \frac{s_d}{s} \quad (6)$$

where q_1 is the rate inclining to the negative emotion, q_2 is the rate inclining to the positive emotion. And $0 \leq q_1 + q_2 \leq 1$, for the emotional inclination of some cybergitizens are not certain. s_0 is the total posts from the appearing time t_0 of a topic to the first moment t_1 of the posts boosting. s is the total posts from t_0 to the first moment t of the posts to the trough. s_n and s_p represent respectively the total posts with negative inclination and positive inclination from t_0 to t_1 . s_d represents the total posts of the opposing view from t_0 to t .

The evolutionary algorithm based on IS&NPEO is as follows:

Input: the cycle index C , N , tr , d_{min} , d_{max} , $[s, s_0, s_n, s_p, s_d]$, $[p, q_1, q_2, k]$.

Output: the opinion set $OP = \{op_1, op_2, \dots, op_N\}$.

Method: do the following steps:

Step 1. According to the input or formula (6), initialize parameters p, q_1, q_2 , and k ;

Step 2. Initialize op_i, e_i, in_i at initial time ($i=1, 2, \dots, N$) as the following rules:

(1) Initialize the op_i of $k*N$ agents as a value in $[0, 0.5]$, and the op_i of other agents as a value in $[0, 1]$.

(2) Initialize the e_i of q_1*N agents and q_2*N agents as a value in $[0, 0.5]$ and $[0.5, 1]$ respectively, and the e_i of other agents as a value in $[0, 1]$.

(3) Initialize the in_i of $p*N$ agents as 1, and the in_i of other agents as 0.

Step 3. Generate the interpersonal network on the basis of the *Small World model*. If there is an edge between agents, they are neighbours, or else they are not neighbours;

Step 4. Select an agent and its neighbors to calculate d_{ij}, I_{ij}, I_{ji} . Then calculate $f(i,j)$, $f(j,i)$ according to formula (3);

Step 5. Calculate $op_i(t+1), op_j(t+1), e_i(t+1), e_j(t+1), in_i(t+1), in_j(t+1)$ after interacting according to formula (4);

Step 6. Each node repeats Step 4 to Step 5 and output the opinion set OP in each loop;

Step 7. Repeat Step 3 to Step 6 until it meets the external termination conditions (completing C loops).

4 Model Simulation and Analysis

To illustrate the rationality and effectiveness of the proposed model, the simulation is implemented on Matlab 2010b. Also this model is compared with the NPO-NW model, in order to explain the influence of the agents' informed state and emotion on the evolution of public opinion.

In the simulation, the number of agents is 100 and the number of cycles is 2000. When the small-world network is generated, the reconnection probability is 0.4. According to the definitions in IS&NPEO model and the results of simulations, the reasonable range of the parameters is: $p \in (0, 0.2]$, $q_1 \in [0, 1]$, $q_2 \in [0, 1 - q_1]$, $k \in [0, 1]$, $op_i \in [0, 1]$, $e_i \in [0, 1]$, $in_i \in \{1, 0\}$. Besides, the ranges of the common parameters owned both by IS&NPEO model and NPO-NW model are the same: $tr \in [0, 1]$, $d_{min} \in [0, 0.5]$, $d_{max} \in [d_{min}, 1]$. In the initial state, op_i and pe_i in NPO-NW model are randomly distributed in $[0, 1]$ and $[0, 0.5]$ respectively. So the experimental results under the same set of parameters are very unstable. Given that when two agents are both informed, the emotional impact factor e in the new model is numerically identical with the uncertainty factor pe model in the NPO-NW. Therefore, in order to ensure the comparison's fairness, and to avoid the influence of random factors to the comparative experiment, the parameters of the NPO-NW model are set as follows: (1) the initial values of $OP = \{op_1, op_2, \dots, op_N\}$ in IS&NPEO model, are identical with the initial OP in NPO-NW model; (2) the initial values of

$E = \{e_1, e_2, \dots, e_N\}$ in IS&NPEO model are divided by 2 to make it meet the requirements of pe in NPO-NW model and set the values to pe ; (3) the values of tr , d_{\min} , d_{\max} in IS&NPEO model are the same as the corresponding parameters in NPO-NW model.

Since k is mainly used for the initial OP , and the initial OP of the two models are the same in the contrast experiment. Hence the initial value of k is 0.5, which means the initial values of half agents are in $[0, 0.5]$ and the others in $[0, 1]$. Table 1 gives the parameters' values in contrast experiments.

Table 1. Parameters of contrast experiments

NO.	p	q_1	q_2	tr	d_{\min}	d_{\max}
(a)	0.01	0.8	0.1	0.5	0.15	0.55
(b)	0.01	0.1	0.8	0.5	0.15	0.55
(c)	0.01	0.4	0.4	0.5	0.15	0.45
(d)	0.2	0.4	0.4	0.5	0.15	0.45
(e)	0.1	0.6	0.2	0.5	0.15	0.45

Figure 1(a)-(d) and Figure 2(a)-(d) respectively give the trends of the public opinion in the new model and NPO-NW model according to the parameters settings in Table1. Figure 1(a)-(c) and Figure 2(a)-(c) reveal the influence of agent s' emotion on the evolution of public opinion when the initial informed ratio ($p = 0.01$) is fixed. They show that agent emotion not only slows down the speed of the evolution of public opinion, but also has a certain impact on the final trend of public opinion;

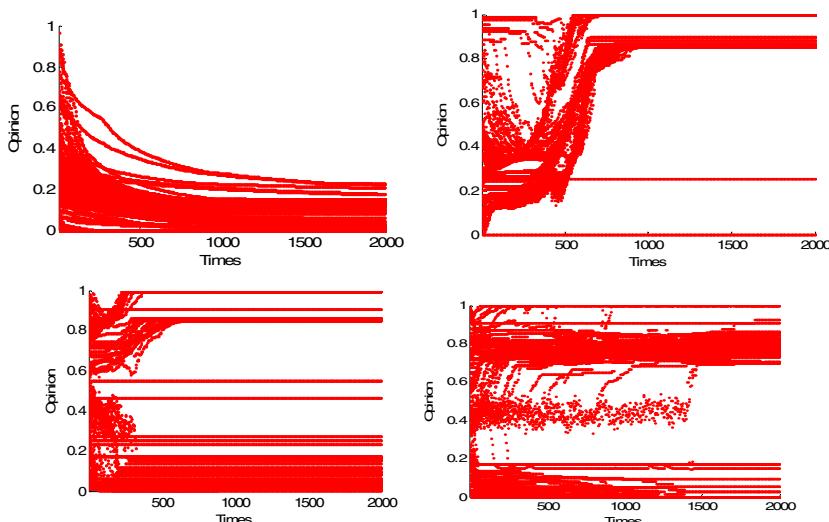


Fig. 1. The evolution of public opinion in IS&NPEO model according to (a)-(d) in Table 1

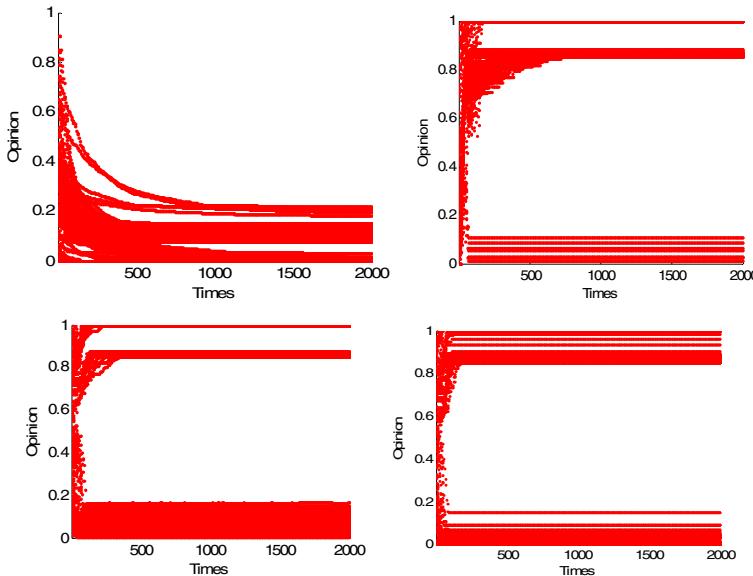


Fig. 2. The evolution of public opinion in NPO-NW model according to (a)-(d) in Table 1

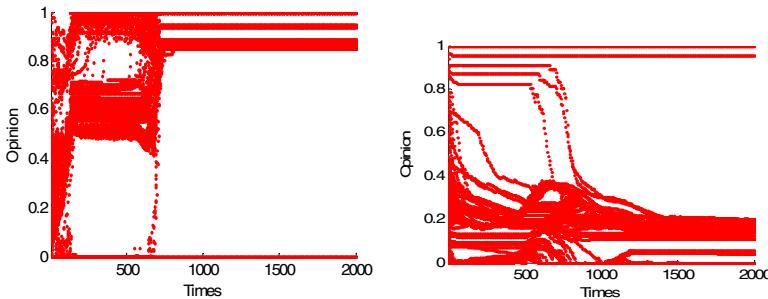


Fig. 3. The evolution of public opinion in IS&NPEO model (*left*) and NPO-NW model (*right*). The parameters correspond to (e) in Table 1.

Figure 1(c)-(d) and Figure 2(c)-(d) reveal the influence of agents' informed state on the evolution of public opinion when the initial emotions inclination ($q_1 = q_2 = 0.4$) is fixed. They show that the agent's informed state has significant influence on the interaction strength, which leads to a remarkably change on the rate of public opinion's formation.

Figure 3 is relatively interesting. It shows that evolution of public opinion exhibited by the new model is almost contrary to that revealed by NPO-NW model. This means that the agent's informed state and emotions have greater effect on the evolution of public opinion, and even may interfere with the final trend of public opinion.

In conclusion, the comparison of our model with NPO-NW model demonstrates that the agent's informed state and emotion affect the rate of public opinion's formation, and interfere with the final trend of public opinion, which accords with the actual interaction of cybecitizens and the formation process of public opinion. The changes of the number of the informed cybecitizens and their emotion definitely exert an influence on the times of the interaction between cybecitizens, and finally affect the formation time and the final trend of the public opinion.

5 Conclusion

An evolution model of emotional Internet public opinion with informed marks is proposed. In this model, the agents' informed status and emotion as well as their initial emotion tendency are considered. Furthermore, the effects of the agent's informed status and emotion on the evolution of public opinion are analyzed.

The experimental results show that agents' informed status and emotion can affect the evolution of public opinion and even change the public opinion's final trend. However, we just make macro-analysis on the pubic opinion's evolution and don't think about other problems such as the changes of network structures. In addition, the external factors in real life aren't taken into account. For instance, the government's policy, the influence of media, war and so on. All of these are our further work.

Acknowledgments. This paper is supported by the following foundations or programs, including National Natural Science Foundation of China(60873079, 61040044), (Key) Natural Science Foundation of Chongqing of China(2009BA2089, 2008BB2241), and Program for New Century Excellent Talents in University (NCET).

References

1. Wu, H., Xin, M.J.: A Quick Emergency Response Model for Micro-blog Public Opinion Crisis Oriented to Mobile Internet Services: Design and Implementation. In: Jin, D., Lin, S. (eds.) 2011 MSEC. AISC, vol. 129, pp. 1–7. Springer, Heidelberg (2011)
2. Liu, C.Y., Hu, X.F., Luo, P., Si, G.Y.: Study on Agent-based Communication Network Model of Public Opinion on Internet. Computer Simulation 26(1), 20–23 (2009)
3. Ding, X.F., Hu, Y., Zhao, W., Wu, R.J., Hu, C.L.: A Study on the Characters of the Public Opinion Leader in Web BBS. Journal of Sichuan University (Engineering Science Edition) 42(2), 145–149 (2010)
4. He, M.H., Zhang, D.M., Wang, H.Y., Li, X.G.: Public Opinion Evolution Model with the Variable Topology Structure Based on Scale Free Network. Acta Physica Sinica 59(8), 5175–5181 (2010)
5. Si, X.M., Liu, Y., Ding, F., Xiong, F.: Research on Bounded Confidence Consensus Emergency Model with Community Structure. Journal of System Simulation 21(23), 7644–7647 (2009)
6. Chen, F.J., Li, L.B.: Application of G (Galam) Model in Network Public Opinion Evolution. Journal of Computer Applications 31(12), 3411–3413 (2011)
7. Sznajd, W.K., Sznajd, J.: Opinion Evolution in Closed Community. International Journal of Modern Physics C 11(6), 1157–1165 (2000)
8. Deffuant, G., Neau, D., Amblard, F., Weisbuch, G.: Mixing Beliefs among Interacting Agents. Advances in Complex Systems 3(4), 87–98 (2000)

Research of LSH and Outliers Detection

Ying-yan Wang, Rui Zeng, Ming-zhong Li, and Fang Li

Yiwu Industrial & Commercial College/ School of Electro-Mechanical

and Information Technology, Yiwu, China

ammy20015@163.com

Abstract. Data supplied mining executive funded data repository of a subset of the sample after the capital data set can replace the original database to reduce order to repeatedly search the database of the time, therefore there is a lot of algorithms has been proposed future at reasonable sampling-owned data sets so that it more real to reflect the original database. These algorithms by data from randomly selected set, select or deletion of swap some noise records of the transactions to make more meaningful rules can be extracted out of the future. We observed that the sample data set is composed of cluster transaction data. Each cluster consists of the similar nature of the information in some of the arguments. Therefore, the removal of outliers should be based on each cluster as a unit without the data set should be based on the entire sample. In order to consider for high-dimensional data encountered curse of dimensionality of the problem. We have studied LSH (Locality the Sensitive the Hashing) the technology to do a cluster of all cut the main point of view through multiple hybrid hash function high similar to the level of trading discipline record will have higher opportunity gathered in the same cluster, the contrary is each other collision reduce the chances.

Keywords: LSH, Outliers, Data Mining.

1 Introduction

Data mining to extract valuable information from large amounts of data into knowledge, that is, the data converted into the knowledge of the behavior. Such information includes general transaction data or multimedia information and knowledge with knowledge is a specific expression and presentation of the implied relationship between the data. Data mining can help companies to obtain the knowledge and create competitive advantage. It caused the majority of the attention, also contributed to the order to the development of new research methods, systems or prototype, and to become a fast-growing field. At present in the data analysis to use on the use of relatively new data mining method, in addition to In order to be information generally funded data repository seemingly useless the conversion into useful information, and to identify hidden the associated nature to excavated out of the relationship between not with the capital data each other, and establish rules for the description of its extent. Data mining technology, frequent item set mining is a very important technology, it can find the contribution data set often appear together, and then dig out the rules of the association rules.

Traditional mining the association rules algorithm must be repeated between the database search as well as to conduct the item sets the merger, in order to find the meet the minimum support high-frequency item set, need a very long time so that efficiency rate is not good. Because today owned data volume the rapid growth of this problem in the future will only be more and more serious. Therefore the design of appropriate data mining algorithms is very important. Future to do association rule data mining by sampling in from a large database is very intuitive way to save the computational burden. In many cases, the approximate results not only sufficient but also satisfactory. Especially in today's changing fast times, fast approximate results than the results obtained in the calculation of the lengthy time-consuming more useful.

Because data mining is a subset of execution in the database, sampling after the data set can replace the original data library to reduce order to repeatedly search the database of the time, therefore there are a lot of algorithms has been proposed future at reasonable sampling of data sets so that it can be more true of the reflect the original data library. These algorithms are selected by randomly selected data set or the deletion of swap some noise records of the transactions to make more meaningful rules can be extracted out of the future.

We observed that the sample data set is composed by the cluster transaction information. Each cluster consists of the similar nature of the information in some of the arguments. Therefore, removal of outliers should be based on each cluster as a unit without the data set should be based on the entire sample. When the cluster is relatively small while the cluster itself also contains some messages, because the comparative lack of representativeness can be treated as outliers and therefore not the same similarity of volume measurement will produce not the same cluster cut, so interest with different perspectives on data set do different cluster classification is necessary. Furthermore, because the clustering algorithm is not absolutely perfect, even if it is assigned to a cluster of information there will be some of the outliers exist, can be for every one of the cluster data to make a further screening. In order to test high-dimensional level-owned data volume will be encountered curse of dimensionality, we adopt LSH (Locality the Sensitive the Hashing) the technology to do the cluster all cut, the main Almost Memorial By multiple hybrid hash function so that high similarity of trading discipline records will be compared to high opportunity to gather in the same cluster, otherwise reduce the chances of mutual collisions.

2 Proposed Scheme

Association rule mining commodity items in the transaction data from a large number of cross, first proposed by Agrawal, after the relevant studies have also been published, our description Ming Guan Stock rules related to the definition next: assume that I is a collection of transaction funded data repository (transaction database) all the items, T set of cross-trading-owned database, all consumers at the turn of trading capital data of a total of m items, each cross-trade-owned data T_j is a collection of some of the items, $i \leq j \leq m$, called the item sets, and $T_j \subseteq I$. An association rule is expressed as $X \Rightarrow Y$ item sets between X and Y , where X is called before the item sets (antecedent), Y called item sets (consequent), $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The association rule $X \Rightarrow Y$ is the effective rule (strong rules) must be determined by two parameters s and c to, respectively, to support level (support) and the trust level

(confidence). Support level s expressed as: in all the transaction data collection (XUY) the ratio value, ie, $s=(\text{contain there (XUY) of transaction data the number of volume})/(\text{all cross-trading-owned data the number of volume})$; while the trust level of c expressed as: contains have X transaction data collection, at the same time also contains have Y the ratio of value, namely $c=(\text{contain there (XUY) of transaction data the number of volume})/(\text{containing an X transaction data the number of volume})$. To mine association rules, the support level and trust level must be greater than or equal to specified minimum support and minimum confidence level, so the association rule makes sense.

Mining association rules is mainly divided into two stages: the first stage, first find out the item sets to meet the minimum support, these satisfy the minimum support the item sets, called frequent item sets. If a item sets contains k items is called k -itemsets, if a k -itemsets meet the minimum support, called the high-frequency k -itemsets. The second stage, with minimal trust level conditions, calculation of high-frequency item sets formed by the association rules, if a minimum confidence level, the associated rules established.

2.1 Apriori Algorithm

In mining association rules, Apriori algorithm is one of the most commonly used method, Apriori algorithm: To finding frequent item sets using candidate of generation. The name of the algorithm is based on the fact: the algorithm will use frequent collection of items the nature of a priori knowledge. Apriori iteration method called layer by layer search, it will use the k -itemsets to explore $(k+1)$ -itemsets. First, we identify the set of frequent 1-itemsets, denoted as the L_1 . L_1 frequently used to find 2-itemset L_2 , L_2 is used to find L_3 , and it goes on until it cannot find frequent k -itemset of the database to do a scan to find each L_k . The algorithm is as follows:

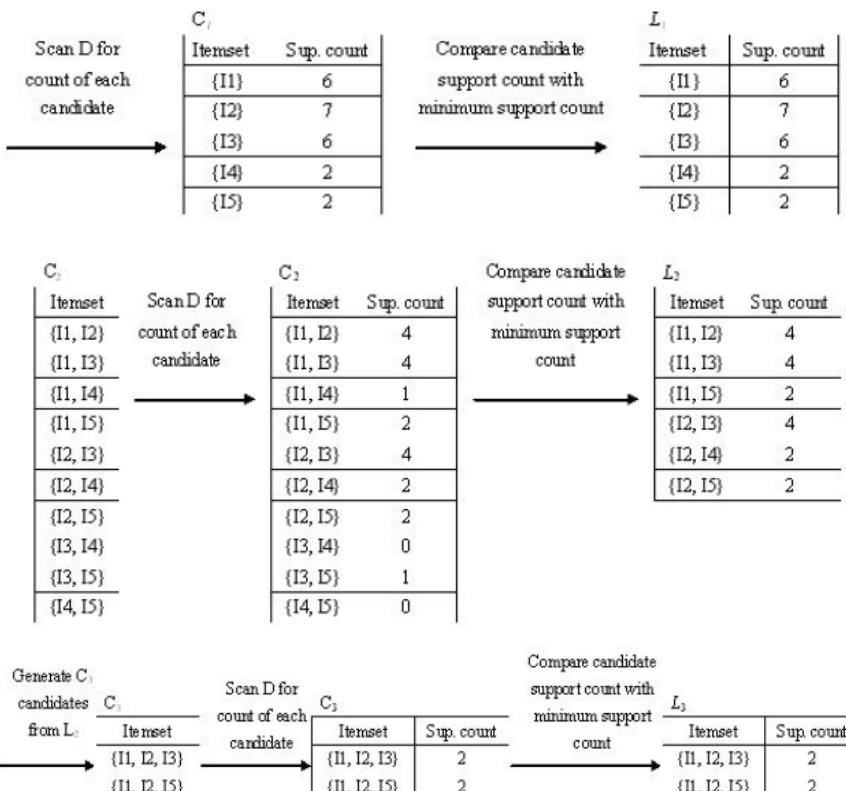
```

 $C_k$ : Candidate itemset of size k
 $L_k$ : frequent itemset of size k
 $L_1 = \{\text{frequent items}\};$ 
for( $k=1; L_k \neq \emptyset; k++$ ) do begin
     $C_{k+1} = \text{candidates generated from } L_k;$ 
    for each transaction  $t$  in database do
        increment the count of all candidates in  $C_{k+1}$  that are contained in  $t$ 
     $L_{k+1} = \text{candidates in } C_{k+1} \text{ with min\_support}$ 
end
return  $\cup_k L_k$ 

```

An example is as follows: suppose only 9 transactions in the database, that is, $|D|=9$. The Apriori assume that the items in the transaction stored in lexicographical order. Explain the Apriori algorithm to find the collection of frequent item in D as shown below:

TID	List of item_IDs
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

**Fig. 1.** An example for Apriori Algorithm

Through the above data mining to find frequent item sets in L_1, L_2, L_3 .

2.2 Two Phase Sampling Based Algorithm

In order to improve the accuracy of the association rules generated by random sampling, propose the FAST-Trim (Finding Associations from the Samples the Transactions) to remove outliers from the initial sample data set to more show characteristics of the original database, so that the association rule mining more accurate. First of all database parent into the behavior extracted sample, and then for the sample-owned data centralized estimate which frequent 1 item set support level, these values will be used to decided which outliers will be from the sample concentrated to remove to get set by sampling the number of volume, Finally to facilitate this final sample set to do the mining of association rules. The algorithm is as follows:

- Obtain a simple random sample S from D
- Compute $f(A; S)$ for each 1-itemset $A \in I_1(S)$
- Using the support computed in Step 2, obtain a reduced sample S_0 from S by trimming away outlier transactions.
- Run a standard association-rule algorithm against to obtain final st. of association rules.

$$\begin{aligned} I_1(T) &= \text{set of all 1-itemsets in transaction set } T \\ L_1(T) &= \text{set of frequent 1-itemsets in transaction set } T \\ f(A; T) &= \text{support of itemset } A \text{ in transaction set } T \end{aligned}$$

The Flow of Fast-Trim Sampling Based Algorithm

In order to decided a cross transaction record for the outliers, need a distance away from the formula as a discriminate criteria, there are a lot of formulas to be set on the future, while the profit with L_1 , the distance away from said is one of the very effective rate of the formula, as follows:

$$Dist_1(S_0, S) = \frac{|L_1(S) - L_1(S_0)| + |L_1(S_0) - L_1(S)|}{|L_1(S_0)| + |L_1(S)|}$$

S is the collection of trade records, the $L_1(S)$ is the number of item sets, which the length of dataset S is 1 frequently, Initial sample for a given data set (S) of the objective is to obtain the final sample data set (S_0), and making the distance among in data set become smallest. Set $S = S_0$, step-by-step operation. First, fast-the trim algorithm data set cut into multiple sizes for a group of k . And for every group, remove $Min(Dist_1(S_0 - \{t\}, S))$ transaction information t make the smallest difference between in the initial sample data set (S) and the final sample data set (S_0). By the order to remove outliers until the set of sample quantity in each group by the final sample data set (S_0). So in order to better streamline and representative sampling of data sets are used instead of the full database to do association rule mining.

3 High Dimensional Data Objects Similarity Search Use LSH

In the space of high-dimensional level, data is relatively sparse in the space, the distance between them is close to a single value, making the expressed distance to outliers meaningless, similarity search become difficult, making the original low-dimensional level of efficient methods become poor, even when the dimension exceeding 10 aging rate is not linear search method, namely, the dimensional level of difficulty phenomenon.

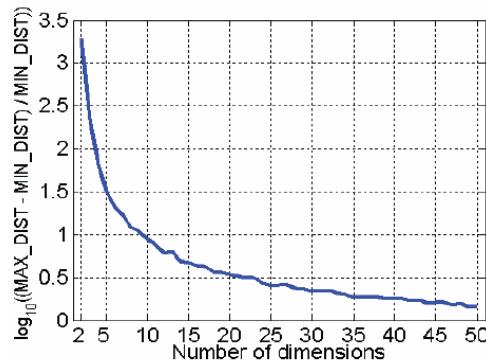


Fig. 2. Curse of dimension

LSH is one kind of very effective rate of clustering methods, the main idea is that each point in space by computing the hash function, can improve the collision probability of each point with similar point. On the contrary, if the distance of two points very far, collision probability will be lower. When inquired, with the hybrid hash function to determine similar points, and obtain data of similar point from the hash group. LSH in the dynamic modification, such as insert, delete, and modify the points, very easy to use, the time complexity is not high. Discussion on LSH, you can use the Euclidean space to calculate the distance between two points, but can also be based on different circumstances, the use of different distance formula to define the different LSH hybrid hash function to determine a similar extent between points. LSH proposed by Indyk and Motwani, whose main purpose is to solve the problem of search of a point in multidimensional space near the point. LSH search for the nearest neighbor point in the space of multi-dimensional level, its time complexity will be less than the complexity of the linear search method.

LSH construction is based on p-stable distribution, each data v use hash function $ha, b(v) = [(a \cdot v + b)/W]$, a is d-dimensional degree of random Gaussian distribution, b is in $[0, W]$ within the uniform distribution, summarize, based on LSH mechanism is by the four participating number of the composition of the W, H, K(specify each a mixed hash table G is the number of hybrid hash function composed of), L(total number of hash table).

- For a d-dimensional vector v , $g(v) = (h_1(v), \dots, h_k(v))$, where $h_j(v) \in H$ for $1 \leq j \leq k$.
- L hash tables are constructed by using g_1, \dots, g_L . Totally, we have L hash tables and a vector v will be hashed into a bucket (indicated by $g_i(v)$) in each hash table.

Mapped from high-dimensional degree of space in low-dimensional degrees of space, all owned data through hybrid hash function mapped to the each interval of W as width to cut into blocks, the following figure shows how the d-dimensional level space on the mapping to a k-dimensional LSH Bucket space linear index space.

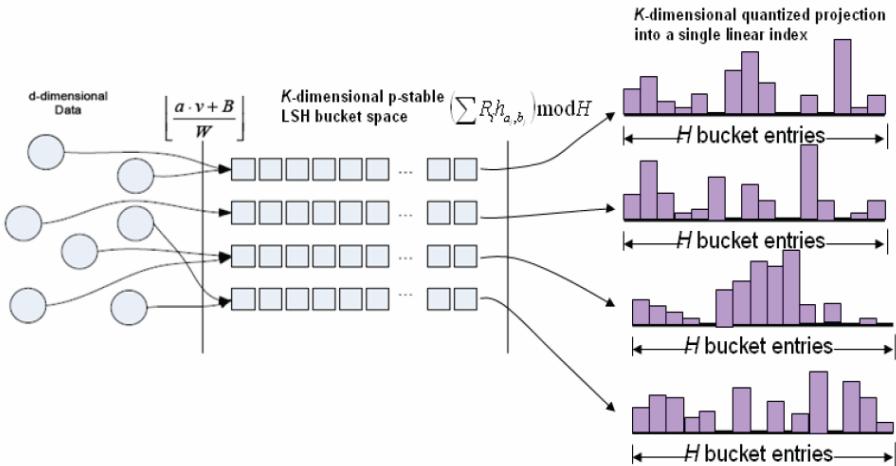


Fig. 3. LSH hash mechanism

4 Conclusion

In the research, we have clear description to data mining association rules, features and other related fields of application, also detected LSH and outliers (outliers) have the solution description.

References

1. Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: The 4th International Conference on Knowledge Discovery and Data Mining (1996)
2. Klein, A., Do, H.-H., Lehner, W.: Representing data quality for streaming and static data. In: The International Workshop on Ambient Intelligence, Media, and Sensing, AIMS Workshop, pp. 3–10 (2007)
3. Fayyad, U.M., Reina, C.A., Bradley, P.S.: Initialization of iterative refinement clustering algorithms. In: The 4th International Conference on Knowledge Discovery and Data Mining, pp. 194–198 (1998)

4. Kraemer, J., Seeger, B.: Pipes - A public infrastructure for processing and exploring streams. In: Weikum, G., et al. (eds.) The 9th International Conference on Management of Data, pp. 925–926. ACM (2004)
5. Guha, S., Rastogi, R., Shim, K.: Cure: An efficient clustering algorithm for large databases. In: The ACM International Conference on Management of Data, pp. 73–84 (1998)
6. Han, J., Kamber, M.: Data Mining: Concepts and Techniques, 2nd edn. Morgan Kaufmann Publishers (2006)
7. Hodge, V.J., Austin, J.: A survey of outlier detection methodologies. *Artif. Intell. Rev.* 22, 85–126 (2004)
8. Mielke, M., Mueller, H., Naumann, F.: Ein data-quality-wettbewerb. *Datenbank-Spektrum* 14, 34–37 (2005)
9. Pérez, J.M., Berlanga, R., Aramburu, M.J., Pedersen, T.B.: A Relevance-Extended Multi-Dimensional Model for a Data Warehouse Contextualized with Documents. In: Proc. Eighth ACM Int'l Workshop Data Warehousing and Olap (Dolap 2005), pp. 19–28 (2005)
10. Nabli, A., Soussi, A., Feki, J., Ben-abdallah, H., Gargouri, F.: Owards an Automatic Data Mart Design. In: 7th International Conference on Enterprise Information Systems (ICEIS 2005), Miami, USA, pp. 226–231 (May 2005)

Research on P2P Botnet Network Behaviors and Modeling

Huabo Li, Guyu Hu, and Yun Yang

Institute of Command Automation, PLAUST Nanjing, China
{huabolee, yyhacker}@163.com, huguyu@189.cn

Abstract. Botnet sustained a serious threat to Internet security. Especially the emergence of P2P botnets, botnet detection has become a very big challenge. This paper focuses on the P2P botnet traffic characteristics and provides support for P2P botnet detection technology. Through a number of experiments, the paper draws some important conclusions, such as high connection failure rate, high outbound network degree, irregular phased-similarity, etc. These conclusions can help the study of P2P botnets detection. The paper also models P2P botnets and proposes a P2P botnet steady-state model. The model can explain some features of P2P botnets are inevitable and these features can be used for more general detection.

Keywords: P2P Botnet, Botnet Modeling, Network Behavior.

1 Introduction

A botnet is a network composed by a large number of compromised hosts (bots) that are remotely controlled by an attacker (botmaster). Botnets can be instructed to commit various malicious activities, such as launching distributed denial-of-service (DDoS) attacks, sending spam, performing click fraud, spreading malicious software, stealing private information, etc. To effectively control a botnet, the botmaster establishes a command and control (C&C) channel with the bots, through which the malicious activities can be coordinated. Botnet has become one of the most serious threats to Internet security. Especially in recent years P2P botnets make botnets more covert, robustness, and dangerous. There are many great challenges in botnet tracking, detection and defense researches.

Network behavior characteristic is the basis of research on traffic based detection of P2P botnet detection. It is believed that the network behavior characteristic of P2P botnet is closely tied to its architecture and operation mechanism. If we illustrate the occurrence inevitability of behavior characteristic by modeling analysis, the specific experiment observations will be extended to general laws and the credibility of the characteristic based detection will be improved.

Our goal is studying the general characteristics of P2P botnets by experimental observations. Then these characteristics can be used to study the common detection methods. To prove or explain these network characteristics observed are really common and right, we want to use mathematical modeling methods to analyze them.

2 Related Work

In recent years, the research on botnets has become an important issue. Zhu et al. [1] divided current research about botnets into three main areas: (1) the investigation of botnets by structural analysis or observing their operation, (2) detecting and tracking botnets, and (3) defending against the attacks of botnets.

Now more and more research focus on botnet detection, particularly P2P botnet detection. There are different types of botnet detection methods, e.g. host-based or network-based systems. The host-based method [2] uses anomaly detection on aggregate network features to identify a deviation from normal activity. Checking the packet payloads being usually needed, it does not work well when many botnets now use encryption to obfuscate their C&C messages. In recent years, more and more detection research use network-based method. [3-9]. These studies have in common is the need to study the behavioral characteristics of the botnets. However, they often do not have systematic. A common understanding of botnets is that multiple (at least two) bots within the same botnet will perform similar or correlated C&C communications and malicious activities. If the botmaster commands each bot individually with a different command/channel, the bots are nothing but some isolated/unrelated infections. There are some modeling studies for botnet. [10-11]. But most of these models focus on the propagation and growth of botnets, rarely related to the network behaviors.

3 Network Behavior Characteristics of P2P Botnet

The process of P2P botnet detection usually includes two steps. First, hosts with P2P characteristic in network traffic are detected. Second, botnet hosts are divided from these hosts. We analyzed the behavior characteristics of normal traffic (without P2P traffic), legal application traffic and P2P botnet traffic, as illustrated in Table 1.

The goal of our experiments is as follows: 1) find the distinction between P2P traffic and normal network traffic. 2) find the distinction between P2P botnet traffic and legal P2P traffic. We measured traffic characteristics such as failed rate of

Table 1. Datasets used in our evaluation

Trace	ID	Duration	# of flows	Description
Storm	D_{storm}	2h	15,915	P2P Botnet, from the Honey Project[12]
Waledac	$D_{waledac}$	10m	12,318	
Thunder	$D_{thunder}$	1h	15,020	File sharing
PPLive	D_{pplive}	1h	7,843	Online video
Skype	D_{skype}	1h	6,100	VoIP
Normal traffic	D_{normal}	1h	2,167	Traffic without P2P
Nmap	D_{nmap}	10m	26,392	Scanning traffic

connection, out-degree of connection and flow similarity, and derived a series of behavior patterns of P2P botnet. Moreover, we analyzed and proved the inevitability of these patterns, and then provided the foundation for P2P botnet detection.

3.1 High Connection Failure Rate

Behavior Pattern I: There are a large number of failed connections in the out-flows of each botnet host.

Due to the churn phenomenon, there are a lots nodes that leaves and joins the P2P network at regular intervals. In order to maintain network connection, P2P nodes send information exchange messages regularly. Therefore, most of out-flows will be failed. This is the common behavior pattern of P2P botnet and legal P2P network.

We define a failed connection as follows: 1) TCP connection is failed if the 3-step handshake is incomplete. 2) UDP connection is failed if there is no response to the query message. As shown in Table 2, the number of failed connection in Storm and Waledac traffic is significantly more than that in normal traffic.

We cut traffic fragments in ten minutes from every trace, then measured the number of failed connection in these fragments. Results are shown in Table 2.

Table 2. Stat. of connection failure rate

Trace	D_{storm}	$D_{waledac}$	$D_{thunder}$	$D_{pptlive}$	D_{skype}	D_{normal}
# of flows	5837	3137	4754	627	442	226
Failed connections	4876	1968	4071	479	422	14
Failure rate	83.54%	62.74%	85.63%	76.40%	95.48%	6.19%

Apparently the fail rate in P2P network traffic is higher than the fail rate in normal traffic. The failed rate of flow can be defined as follows:

Definition 1: Connection Failure Rate is the radio of the number of failed connection with the number of out-flows in a predetermined time period. The number of failed connections contains failed TCP connections and UDP connections.

Fig 3 shows the failed rate of Storm, Waledac and normal traffic. It is obvious that the failed rate of botnet traffic is higher than that of normal traffic. We call this behavior pattern as High Connection Failure Rate (HCFR), and HCFR is a critical parameter to detect P2P botnet traffic.

3.2 High Outbound Network Degree

Behavior Pattern II: In the active period, host of P2P botnet will connect with a large numer of nodes distributed in different networks.

Definition 2: Outbound IP-Degree (IPD) is the number of outbound addresses which a host accesses with in a predetermined time period.

High outbound IP-degree is an important characteristic of a P2P bot. However, some normal network applications also have this characteristic. For example, Outbound IP-Degree and failed rate of network scanning is extremely high. This will produce adverse effects on botnet detection by outbound IP-Degree. To improve the accuracy of botnet detection, we define Outbound Network-degree as follows:

Definition 3: Outbound Network-Degree (OND) is the number of outbound networks which a host accesses in a predetermined time period.

In our experiments, we identified the accessed network of by deriving the BGP prefix of IP addresses. [13]. We cut traffic fragments in ten minutes from the traces, then measured IPD and OND in these fragments. Results are shown in Table 3.

Table 3. Stat. of outbound network-degree

Trace	D_{storm}	$D_{waledac}$	$D_{thunder}$	D_{pplive}	D_{skype}	D_{nmap}	D_{normal}
IPD	5561	3111	3642	233	54	6798	53
OND	2357	2233	1685	190	53	4	39
OND/min	236	223	169	19	5	1	4

In Table 3, both the IPD and OND of P2P traffic are high. The IPD of D_{nmap} is high, while the OND of D_{nmap} is low. Both the IPD and OND of normal traffics are low. Therefore, we will adopt High Outbound Network Degree (HOND) as behavior pattern to detect P2P botnet traffic.

3.3 Irregular Phased-Similarity

Behavior Pattern III: P2P botnet flows have similar patterns (outbound network degree, average packets number of flows, connection failure rate, etc.), which occur at irregular intervals.

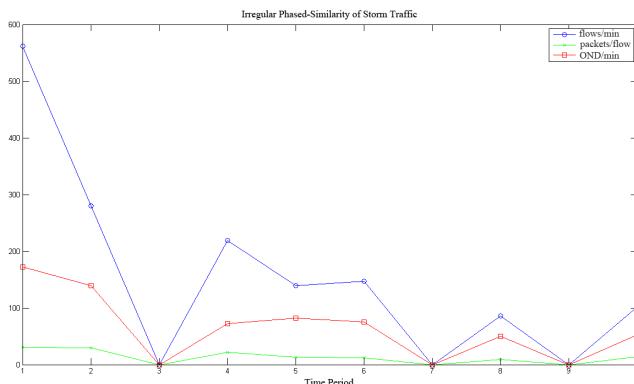


Fig. 1. Irregular Phased-Similarity of Storm traffic

For maintaining the network, every node in the P2P botnet needs to send query messages to check whether the neighbor nodes are reachable. Some obfuscate technologies are used by the botmaster to restrain the regularity of nodes discovery. So the periodic messages appear at irregular intervals. We call this Irregular Phased-Similarity (IPS).

Take Storm traffic for example, we count the flows number, packets number per flow and outbound network degrees on the trace, and the IPS law is very clear, as show in Fig.1.

IPS is an important network feature of P2P botnet, which can be utilized to separate P2P botnet traffic from legal P2P network.

4 Modeling P2P Botnet

Based on the research on the architecture and operation mechanism of P2P botnet, we build a mathematical model for the DHT based P2P botnet to analyze the behavior patterns of P2P botnet.

4.1 P2P Botnet Steady-State Model

First, we define the operation mechanism and relevant parameters as follows:

Suppose the number of nodes in the P2P botnet is M , and each node maintains a routing table, which is also called neighbor node list. The elements in the list are the IP addresses of reachable nodes, and we refer to the number of these elements as node degree. The design of node degree d varies according to the structure of P2P protocol (such as Chord, CAN), and we let $d = nk$ where n is the value space of node ID.

Each node send query message at regular intervals to check whether the neighbor nodes are reachable. Let Δt as the predetermined time period, each node exchange routing table with its adjacent nodes, and then update routing table according to the received node lists. Δt is also called as update cycle of routing table.

We assume that the population of the P2P botnet is constant when the network reaches a steady state. In this state, we let μ as the exiting probability of the nodes within time period Δt . It is obviously that μ is also the adding rate of new nodes.

Assume α as the intersection rate of IP addresses sets between two adjacent nodes, it is clear that the average number of repeated IP addresses of two adjacent nodes is $d\alpha$.

Consider the node P_i in the P2P botnet, the adjacent node of P_i leaves with a probability μ at regular interval Δt , and the number of reachable adjacent nodes reduces to $d(1-\mu)$. Then P_i exchanges the routing table with the remaining adjacent nodes, and the number of reachable node of P_i will be revert to d .

In the update process of node list, each adjacent node will provide $d(1-\alpha)$ nodes to P_i . Therefore, the number of updated nodes provided by the remaining $d(1-\mu)$ nodes is $d(1-\alpha) \times d(1-\mu)$. Considering the redundancy between nodes is $d+1$, the number of updated nodes is:

$$d(1-\alpha)d(1-\mu)\frac{1}{d+1} = \frac{d^2}{d+1}(1-\alpha)(1-\mu)$$

The number of the adjacent nodes will be updated into a balanced state. In other words, the degree of node will remain unchanged. So:

$$d(1-\mu) + \frac{d^2}{d+1}(1-\alpha)(1-\mu) = d \quad (1)$$

We call the above equation as ‘equation of P2P botnet in steady state’, and therefore deduce that:

$$\alpha = 1 - \frac{(d+1)\mu}{d(1-\mu)} \quad (2)$$

When the $d=2^n k$ is great enough, $\frac{d}{d+1} \approx 1$, so:

$$\alpha \approx \frac{1-2\mu}{1-\mu} \quad (3)$$

We call α as the robust parameter of P2P botnet, and the above formula is called botnet robustness function. According to formula (3), the robust parameter decreases drastically as the node exiting probability increases. Apparently the condition $\mu \in (0, \frac{1}{2})$ should be satisfied to construct a steady P2P botnet.

Based on the above analysis, we propose a P2P Botnet Steady State Model (P2PB-SSM) as follows:

(1) When $\mu \rightarrow 0.5$, $\alpha \rightarrow 0$. In the update cycle of node list, if the node exiting probability μ close to 0.5, the intersection set of the adjacent node lists between neighbor nodes is almost empty. In this case, all nodes cannot derive new nodes from their adjacent nodes and will leave from the botnet.

(2) When $\mu \rightarrow 0$, $\alpha \rightarrow 1$. In the update cycle of node list, all nodes will be maintained in the botnet, and the intersection set of the adjacent node lists between neighbor nodes will reach maximum scale. In this case, the robustness and toughness will be maximized.

According to P2PB-SSM, the following theorem is obtained:

Theorem 1: In the update cycle of routing table, the exiting probability μ should not exceed 0.5.

4.2 Behavior Characteristics and the Model

If the degree of the botnet node we limit, P2PB-SSM is also suitable to describe botnets that using other protocol. In practice, the degree of botnet node is restricted because of limited network bandwidth.

Based on the framework of P2PB-SSM, we focus on the routing table update period of a botnet node. Let μ_0 be the exiting probability of time period Δt , the probability distribution of node exiting can be defined as follows:

$$\mu(t) = 1 - (1 - \mu_0)^{\frac{t}{\Delta t}} \quad (4)$$

Let τ be the maximum value of neighbor nodes update period of the P2P botnet. P2PB-SSM shows that the exiting probability μ will be 0.5 when τ is used as the update period. In this situation, the botnet reach a critical state that the network cannot be constructed, as shown in Fig. 5. According to formula (3):

$$\mu(\tau) = 1 - (1 - \mu_0)^{\frac{\tau}{\Delta t}} = \frac{1}{2}.$$

It is obvious that:

$$\tau = \frac{-\log 2}{\log(1 - \mu_0)} \Delta t \quad (5)$$

τ is also named as the maximum designed period of node list update. If the update period exceeds τ , the botnet will not be constructed. Therefore, the range of update period in botnet design process is $(0, \frac{-\log 2}{\log(1 - \mu_0)} \Delta t)$.

Combined with P2PB-SSM, we can conclude the following deduction:

Deduction 1: The node exiting probability being unchanged, the robustness of the P2P botnet will be improved as the update period of routing table decreases.

P2PB-SSM model can be used to assist in the analysis of P2P botnet behavior characteristics. We can see that some phenomenon in the experiment was not accidental.

For example, high connection failure rate in the P2P botnet traffic is bound to exist. According to the formula (4), the exiting probability of a node is $\mu' = 1 - (1 - \mu_0)^{\frac{t'}{\Delta t}}$. If the neighbor nodes number is d , there are $d\mu'$ nodes will no longer in the adjacent node list. In this case, the connection failure rate will be higher than μ' . Considering other abnormal factors in practice, the rate will be much more higher. The maximum update cycle of P2P botnet is τ . This means that, regardless of the botmaster is how to design a P2P botnet, the average update period can not be longer than τ . So in a P2P botnet, the node must refresh the routing table after a certain time. Then we know that the IPS law there will be.

5 Conclusion and Future Work

In the paper, we analyzed the P2P botnet traffic characteristics, such as high connection failure rate, high outbound network degree, irregular phased-similarity,

etc. We also model P2P botnets and proposed P2P botnet steady-state model and maximum update cycle model. These models can explain some features of P2P botnets are inevitable and these features can be used for more general detection.

In future works we will perfect the P2P botnet models, which should be much more common and realistic. Based on the network behavior characteristics and models researches, we will study general P2P botnet detection technology to detect known and unknown P2P botnets.

References

1. Zhu, Z., Lu, G., Chen, Y., et al.: Botnet Research Survey. In: 32nd Annual IEEE International Computer Software and Applications Conference, Turku, Finland, pp. 967–972 (July 2008)
2. Wurzinger, P., Bilge, L.: Automatically Generating Models for Botnet Detection. In: European Symposium on Research in Computer Security (2009)
3. Gu, G., Perdisci, R., Zhang, J., et al.: BotMiner: clustering analysis of network traffic for protocol- and structure-independent botnet detection. In: SS 2008: Proceedings of the 17th Conference on Security Symposium, pp. 139–154. USENIX Association (2008)
4. Ruben, D., Torres, M.Y.H.S.: Inferring undesirable behavior from P2P traffic analysis. In: SIGMETRICS 2009: Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems. ACM (2009)
5. Su, C.: T E D. P2P botnet detection using behavior clustering & statistical tests. In: AISec 2009: Proceedings of the 2nd ACM Workshop on Security and Artificial Intelligence. ACM (2009)
6. Noh, S., Oh, J., Lee, J., et al.: Detecting P2P Botnets Using a Multi-phased Flow Model. In: ICDS 2009: Proceedings of the 2009 Third International Conference on Digital Society, pp. 247–253 (2009)
7. Yen, T., Reiter, M.K.: Are Your Hosts Trading or Plotting? Telling P2P File-Sharing and Bots Apart. In: ICDCS 2010: Proceedings of the 2010 IEEE 30th International Conference on Distributed Computing Systems, pp. 241–252 (2010)
8. Yen, T.: Detecting Stealthy Malware Using Behavioral Features in Network Traffic. Carnegie Mellon University Department of Electrical and Computer Engineering (2011)
9. Zhang, J., Perdisci, R., Lee, W., et al.: Detecting stealthy P2P botnets using statistical traffic fingerprints. In: DSN 2011: Proceedings of the 2011 IEEE/IFIP 41st International Conference on Dependable Systems & Networks, pp. 121–132 (2011)
10. Ruitenbeek, E.V., Sanders, W.H.: Modeling Peer-to-Peer Botnets. In: QEST 2008: Proceedings of the 2008 Fifth International Conference on Quantitative Evaluation of Systems, pp. 307–316 (2008)
11. Langin, C., Zhou, H., Rahimi, S., et al.: A self-organizing map and its modeling for discovering malignant network traffic. In: CICS 2009: Computational Intelligence in Cyber Security, pp. 122–129 (2009)
12. French Chapter of the Honeynet Project,
<http://www.honeynet.org/chapters/france>
13. <http://www.team-cymru.org/Services/ip-to-asn.html>

Web-Based Solution for the Diversity of Equipment

Chang Yan-shuo and Hua Qing-yi

Dept. of Information Science and Technology, Northwest University, 710127 Xi'An, China
changyanshuo@foxmail.com, huacy@nwu.edu.cn

Abstract. With various carriers which transmit information arising, they further meet the needs of users. However, they also bring some new problems, such as difficulties in the data communication among devices, data inconsistencies in devices, complexities with application development, etc. This paper presents a new network structure, which uses the current existing and mature technology to solve the problems caused by hardware diversity. Using the new network structure, we can reduce data redundancy and data inconsistency, making the distributed software own a higher availability, reliability and better maintainability.

Keywords: Network Structure, Availability, Reliability, Data Communication.

1 Introduction

Information is an object [1], which supplies a more convenient way for people to access and transmit civilization created by human, promotes and enriches interpersonal communications among people. At present, it has played a major role in social life. There is no doubt that modern society is information society.

Information has such traits as: transformability and transitivity. Transformability means information can be changed or converted by processing it to change form and content for meeting special needs [2]. Information transitivity means the storage of time, transfer or diffusion means the storage of space. People have developed kinds of personal equipments using these traits, for example: PC (personal computer), e-ink book, phone (Smart phone), PDA (Personal Digital Assistant), MP3(Moving Picture Experts Group Audio Layer III) player, PMP(Portable Media Player), PMC (Portable Media Center), PVR (Personal video Recorder), PMA (Pocket Media Assistant), PSP (PlayStation Portable), etc.

Equipment diversity caters to the personal needs of people, which can use different equipments to view or deal with information at different times, places and moods conveniently. It also results in a number of new problems [3-6].

The data redundancy and inconsistencies are the most urgent problems [5, 6], the devices with information are amounting, and data synchronization between them are further cumbersome, this not only causes data redundancy, but also prone to confusion and inconsistency. For example, a person downloads a book from the Internet by his office computer, and he still hasn't finished the book after work, so he needs to copy this book with annotation to his electronic paper book. The next day, he

needs to synchronize this annotation with PC. This makes people irritable and a slight negligence will cause data inconsistencies between PC and electronic paper book. This example only involves two devices. If using a lot of equipment, he will face many synchronistical operations. This heightens the burden to the user and increases the possibility of data inconsistency.

Different users need to synchronize data with different devices, and different hardware has installed individual software, and the software also needs to maintain by users. But these are not the necessities which users have to do, which reduces the software availability.

Can we find a good network structure to solve this problem? Section 2 presents a network structure, and how to achieve its application will be discussed in section 3, section 4 presents the conclusion.

2 The New Network Structure

2.1 Logical Structure

Logical structure of the new is shown as below, which consists of three levels: presentation layer, logic layer and storage layer. Presentation layer corresponds to interface which people can look and feel, the main function of this layer is to interact with user through the use of interactive browser, which is a presentation of information users concern about in a certain way and it provides users with flexible and convenient ways to interact. Logic layer corresponds to the logical process, information, data, the existence and levels of this information as well as the relationship between the information. Storage layer is the physical storage structure, which adopts network storage technology to store the information and data with the ability of fault tolerance, transparency and security.

The main features of this logical structure are the separations between the presentation layer and logic layer, storage layer and logic layer.

The related aspects of information are information storage, information processing and information presentation which users especially concerned. They don't want to know the software version in details, from where to read the data, and how they get final state of information. Therefore, shielding the details of information processing and information storage to make its transparent to users can improve system availability.

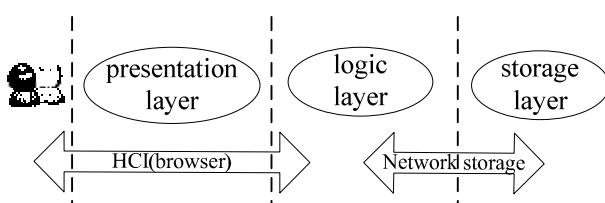


Fig. 1. Logical structure

2.2 Physical Structure

A user usually has one or more physical devices, one of them owned by one or more users. The program and data in previous physical structure are separately stored in different physical devices. Users take advantage of different physical devices to interact, access information, and update the data or programs on these devices. Information is copied and synchronized between different physical devices through wired or wireless internet network.

On the contrary, users interact with a variety of physical devices in the new physical structure. The main interactive way is that devices present the information through the browser, and users access the program which provided by the program provider in the internet, and then operate the stored data by corresponding programs. Sometimes users can directly access the information stored on the internet through the utilization of browser, update the stored information through a browser after use, edit and feedback.

There is a major difference towards previous physical structure—for the same logical file or application, the previous structure will be in the form of multiple backups on different devices while the new structure has only one physical file which stored in the network.

2.3 Key Technologies and Influencing Factors

Being confined with transmission speed, storage and other technological, previous network can't achieve this structure, but now it is critical to implement this structure with the development of high-speed transmission and successful application of network storage. The key technologies of network structure are network storage, network bandwidth and web service.

2.3.1 Network Storage

NAS (Network Attached Storage) [7,8] means storage devices connect to a group of computers by standard network topology (e.g., Ethernet). SAN (Storage Area Network) [9] connects to a group of computers by Fiber Channel. Network storage technology successfully achieves the transparent storage, and it's easy to implement data disaster recovery, while the application programmers generally needn't know its detailed storage.

From Google's practice [10], they use a scalable distributed file system Google File System (GFS), which is an efficient storage technology cluster. With the development of clustered storage technology, it not only breaks easily the SAN bottlenecks, but also implements the linear expansion of performance and capacity. It's a new choice for an enterprise in pursuit of high performance and availability.

IBM recently has released the solution of newest cloud storage—IBM Smart Business Storage, and declared: "This program completely support the protocol of standard file access, making the data migration to the cloud or move out from it as simple as performing file copy operation".

Symantec has also introduced a cloud environment for the cluster file system File Store, and it said the system can provide 40PB online storage space for 900 million users.

2.3.2 Network Transmission Speed and Security

The current transmission speed becomes faster and faster, next-generation transmission system with 40Gbps has been commercially used. Anaheim located in California, held Optical Fiber Communication Conference and Exhibition on Nov 2010. Lucent Technologies Bell Labs has announced the first success in the global electronic data multiplexed 107Gbps optical transmission, and 160Gbps optical fiber communication systems have also been successful in laboratory experiments.

2.3.3 Other Possible Related Factors

In addition to Network Storage and Network transmission speed and security, there are plenty of other possible related factors that affect the network model's implementation, such as the governments' policies and the industries' (especially big companies) support.

This model includes the network infrastructures, software developers, providers, defenders and end users, which may bring about the failure of the whole network structure resulting from any incongruity of every single process. Because of the small enterprises and free software developers' incapacity to promote the implementation of the model, this can only rely on the governments' impetus and big businesses' guidance.

2.4 Advantage of the New Model

2.4.1 Low-Cost

The data is stored on the network instead of terminal equipment, one advantage of this is that the efficiency of disk utilization can be improved so that decrease the hardware cost of entire system. At the same time, because the software itself is only necessary to be existent in the network, the users do not need to install and maintain the software, which, therefore, saves their time and reduces their maintenance costs.

2.4.2 Improve the Reliability of the Software

Because the user' data takes the network storage and they just check through and use the logical view of the data with HCI browser, the data therefore only saves one copy. Just as the database theory, saving one copy of the data reduces the data's inconsistencies. Therefore, the proposed architecture improves the reliability of the software.

2.4.3 Improve the Usability of the Software

Availability sheds light on a set of properties related to individual appraises which combine efforts with a group of disciplines or implied users. It mainly concludes: understandability, learnability, operability.

The new network structure has following advantages: 1. Processing software for each device without having to install; 2. Does not require users to manually synchronize the data, thus users can focus more on the business than on the detail of technology; 3. Users do not need to consider upgrading the software or patch for software. They need the results rather than process, and they don't want to know too much detail. The internal implementation details are transparent to the user, and the method reduces the burden of user increases the user's ease.

Since different browsers can provide a different view of user interaction, so it can better meet the various needs of different types of users, it's an electronic book which can be presented by their favorite font, black and white style in electronic paper, MP3 can use directly TTL to read for users. It also provides visual interaction, auditory interaction, sensory interaction or its synthesis, effectively improving the ability of understanding and learning.

2.4.4 Improve the Maintainability of the Software

Due to the separation of logic layer and presentation layer, if users want to change needs, programmers only change the presentation layer instead of the contents of the storage layer and logic layer, which greatly improves the maintainability of the software. Because some software services require real-time use, such as bank teller machine systems or mobile communication systems, interrupting the services to upgrade and maintain the software means huge economic losses. The solution proposed in this paper can also resolve this issue by relocation of the services.

3 Application Examples

3.1 3A Mobile Office

Mobile office is also called 3A office. Office workers may at anytime, anywhere processing anything with business. Using this model on mobile office business, for example, Mr. Wang uses the computer to edit a document in the office, and then his wife called him to have dinner after work, but, at the same time, this document must be finished today, so he can directly save it, open this document on the subway, it is negligible to consider all the synchronous work and worry about the consistency of document.

3.2 Online Document Editing

Online document editing is becoming increasingly popular [11]. Compared to traditional document editing software, the appearance of Google Docs will dramatically change the way and our habit of using. we will no longer need to install game software and other office software on our PC, just open the Google Docs web to edit and update documents (Using cloud computing systems), store the documents in

personal storage space (Using cloud storage systems). No matter where we go, we can open the saved documents in the cloud storage system after logging on Google Docs. It's easily to share, transmit, and manage copy through the rights management in cloud storage system.

Following the Google Inc., Microsoft has introduced online documentation service named docs, and the docs' most prominent feature is both the program and data stored in remote network.

4 Conclusion

This network has become a trend, and some software companies have released application. Although previously cited Kingsoft's WPS, Google docs and Microsoft docs, etc. have achieved this structure more or less, but not every person or company have recognized that. This paper hopes that clear this architecture in theory, system engineers can design good conception models [12], then development network software with it. We should accelerate this transition between the old and new architecture, making this transition from spontaneous into self-conscious.

Acknowledgments. This work was supported by NWU Graduate Cross-discipline Funds (10YJC06), the Shaanxi Provincial Department of Education research project (2010JK850) and NSFC (60873095).

References

1. Kaldoudi, E., Dovrolis, N., Dietze, S.: Information organization on the internet based on heterogeneous social networks. In: The 29th ACM International Conference on Design of Communication, Pisa, Italy (October 2011)
2. Kim, Y.-W., Yoon, K.-B., Woo, Y.-G., Woo, J.-H.: An Application of Multimedia Teaching Lesson for Children with Developmental Disabilities. Routledge, London and New York (1995)
3. Wright, A.: Get Smart. Communications of the ACM 52(1), 15–16 (2009)
4. Kovachev, D., Renzel, D., Klamma, R., Cao, Y.: Mobile Community Cloud Computing: Emerges and Evolves. In: Proceedings of the First International Workshop on Mobile Cloud Computing (MCC). IEEE, Kansas City (2010)
5. Barra, A., Agliari, E.: A statistical mechanics approach to Granovetter theory. Statistical Mechanics and its Applications 391(10), 3017–3026 (2012)
6. Motro, A., Anokhin, P.: Fusionplex: Resolution of data inconsistencies in the integration of heterogeneous information sources. Information Fusion 7(2), 176–196 (2006)
7. Gibson, G.A., Van Meter, R.: Network attached storage architecture. Communications of the ACM 43(11), 37–45 (2000)
8. Zhang, Y., Zhou, Y.: Separating computation and storage with storage virtualization. Computer Communications, 1539–1548 (August 2011)

9. Ohtani, A., Aono, H., Tomaru, H.: A File Sharing Method for Storage Area Network and its Performance Verification. *NEC Research & Development* 44(1), 85–90 (2003)
10. Dekeyser, S., Watson, R.: Extending Google Docs to Collaborate on Research Papers. Technical report, The University of Southern Queensland, Australia (2006)
11. Farkas, D.K., Poltrack, S.E.: Online editing, mark-up models, and the workplace lives of editors and writers. *IEEE Transactions on Professional Communication* 38(2), 110–117 (1995)
12. Wilkinson, K., Simitsis, A.: Designing Integration Flows Using Hypercubes. In: *EDBT* (2011)

Fusion Assessment Methods for Bridge Health State Based on Two Step Neural Networks Ensemble

Li Shan¹, Zhao Hui², and Rao Hong¹

¹ The Center of Computer, Nanchang University, Jiangxi Province, China

² The Computer Science Department, Nanchang University, Jiangxi Province, China

{lishan, raohong}@ncu.edu.cn, 344913345@qq.com

Abstract. There are multitudinous parameters during the processing of monitoring the bridge, which leads to difficulty in predicting the state of bridges. Thus, it is difficult to acquire the bridge health state accurately according to traditional methods. In order to coalesce different types of monitoring parameters or the asynchronous data monitored by the same type parameters, and obtain the concordance evaluation of the bridge health state, this paper presents a fusion assessment method for the bridge health state based on two step neural networks ensemble. The method reduces the complexity of process on the multisource fusion of the monitoring data and improves the accuracy of the bridge health state assessment.

Keywords: Two Step Neural Networks Ensemble, information fused, Bridge Health Assessment.

1 Introduction

Neural network integration is a very effective engineering neural calculation method, which was put forward by Salomon and Hansen first in 1990. They had proved that the generalization ability of neural network could be significantly improved by training multiple neural networks and fusing the result of each network. The method can be used in engineering neural calculation effectively.

The bridge health monitoring involves multitudinous parameters, it could not give a precision assessment by single neural network. Besides that, the paper proposes a fused method which is used to assess the bridge health state by neural network integration. Firstly, we used the minimize error for classifying the standard to choose the best output network, and then estimated the optimal linear power to individual network integrated to form ideal classification. We introduced the method that takes use of neural network for training the different types of signal collected by the displacement sensor, stress sensor, environmental testing related sensor etc. and the system gives the assessment by training results integration. The method could give guidance for the maintenance of the bridge.

2 Basic Thought of Neural Networks Ensemble

2.1 Conception of Neural Networks Ensemble

Solich and Krogh defined the neural networks ensemble as neural networks ensemble with the limited is a neural network to the same question for learning, and integrating a sample of input output by an integrated the neural network in this example of output decided together. However, some researchers think that neural networks ensemble refers to multiple independent training neural network learning and codetermine the finally output which is not required to learn the same question in the ensemble networks.

2.2 Process of Neural Networks Ensemble

In this paper, when classified by the neural networks integration methods, firstly using minimize error for standard classification chose the relative to each output to the classification of the best network, and then estimated the optimal linear power to individual network integrated form ideal classification.

2.3 Generalization Ability Analysis of the Neural Networks Ensemble

Suppose the integration consist of N independent neural network classifiers, and take the absolute majority ballot method, and also each network could give the correct classified result by the probability of $1-p$, at the same time, we suppose that the mistakes are not association, the probability of the error from the neural network is:

$$p_{err} = \sum_{k>\frac{N}{2}}^N \binom{N}{k} p^k (1-p)^{N-k} \quad (1)$$

When $p < 1/2$, p_{err} will be montone decreasing as N increasing. Consequently, if every calculate accuracy of the neural networks ensemble overtop 50%,and the mistakes between networks are not related to each other ,the more neural networks ensemble, the higher the accuracy. When N tend to infinite, the integrated error rates tend to 0.

In the practical applications, As a result of each independent neural network does not ensure that mistake is not relevant, therefore, we use weighted average in order to improve the generalization ability of the neural networks ensemble. Proper selection of each network weights, get more than a simple method of mean better generalization ability.

Suppose the mission of study is that use a neural network of integrated to $f : R^n \rightarrow R$, and integration using the weighted average, the generalization error of the neural networks ensemble is as follows:

$$E = \bar{E} - \bar{A} \quad (2)$$

3 The Parameter Selection for Assessment

The bridge health assessment technology is mainly through to the real-time monitoring of each structure of the bridge, so that it could be used in the early warning work for the unexpected emergency. The sensors would not only used in the real-time data monitoring of the local bridge components, but also the other environmental factors which beyond the bridge itself. With the data, the system will give the prediction for the bridge life.

According to the structural characteristics of the bridge, the monitoring part of the bridge comprised by the following two categories:

(1) structural static monitoring

To the general bridges, the structural static monitoring includes displacement monitoring, dynamic load acceleration monitoring, local deformation parameters monitoring.

(2) other variable monitoring

Other variable monitoring includes the surrounding environment wind speed, wind load monitoring, the environment temperature monitoring, traffic load and the seismic load, the boat struck load monitoring, etc.

From analysis above, in order to obtain the bridge health state, it needs to monitor different types of parameters. And also for the same parameters, it needs to monitor in the different key point of the bridge. Therefore, complex monitoring data may influence by each other, and through the traditional neural network methods, it is difficult to obtain the bridge health state. So in this paper , it puts forward the bridge health state assessment which based on neural network ensemble fusion assessment method, to integrate different types of monitoring parameters and the parameters of the same type multi-point asynchronous data, obtain the bridge health state consistency assessment.

4 The Assessment for Bridge Health Based on Neural Networks Ensemble

4.1 The Method for Assessment

The assessment of the bridge health is made of 3 parts: the data from transmission sensors, the operation on the system by humans, the knowledge of the bridge health assessment. It's so difficult to find the inner relationships between the key parameters of the bridge that we can't make the exact assessment of the bridge health state.

We use the two step neural networks ensemble method to realize the assessment of the bridge health state. The N- types parameters which are monitored are input into several integrations, and each integration is given a credibility value, and it is as the weighting index in the second step ensemble, and combines each integrations by the weighting index. Each single neural network pre-learns the different types health data, the construction problem data, the construction problem indications which are related

with the bridge's conditions, and finishes a kind of linear mapping relationship to make an comprehensive assessment of the bridge's health. This method not only solve the problem that the neural network works slowly when a large quantities of data was needed, but also make the assessment of the bridge health state be much easier.

4.2 Construction of the Two Step Neural Networks Ensemble

We constructed the two step neural networks ensemble for the assessment of the bridge's health by the method and the research in above chapter.

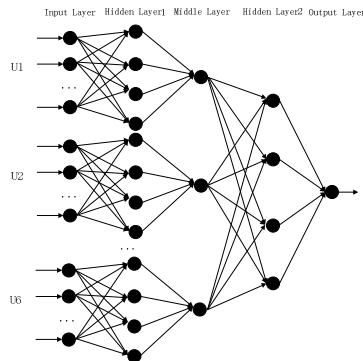


Fig. 1. The structure Two Step Neural Networks Ensemble

(1)Groups of Attribute

U1 (monitoring parameter1)= {U11(key part1), U12(key part 2), U13(key part 3).....}

U2 (monitoring parameter2)= {U21(key part1), U22(key part2), U23(key part3).....}

.....

(2)Groups of Result

We divide the bridge health states into 5 types according the actual demand, and make a group of results.

$$V=\{v1(\text{good}), v2(\text{better}), v3(\text{middle}), v4(\text{bad}), v5(\text{worse})\}$$

4.3 Design of the Sub-nerve Network

For each child attribute set of the neural network is designed by the two layers of BP neural network structure. Two layers of BP neural network can easily achieve nonlinear hint. According to each component of the attributes of the key parts of the corresponding to the number of node determine inputs, each key part sets a sensor and referred to a node in the input layer. Output layer corresponds the result set, set the output node for just one. And according to experience formula $n=\sqrt{(ni+no)}+C$, chosen different numbers of hidden layer node.

With bridge deformation as an example, Wushahezha traffic bridge set 18 deformation monitoring stations, the values of hidden nodes should be between 5 to 15. Testing, when take 9 hidden nodes for the best. In another view, the key testing parameters including wind load, temperature load, traffic load, displacement, dynamic load acceleration and deformation etc., hiding the node values range should be from 3 to 13, so taking 6 hidden nodes should be best.

According to the requirements and characteristics of bridge health monitoring, it is of high accuracy for the output. When it contains hundreds of weights, the L-M arithmetic converged fastest, and the transmission function using the LOGSIG function.

5 Simulation Test

Wushahezha traffic bridge which is placed in the north part of the city of Nan Chang, a fabricated prestressed concrete hollow slab bridge, consists of 5 apertures, 14m span, bridge deck width is 1m+13m+1m, that the design load level is:car-20, trailer-100. We analysis the bridge structure by a bridge structure analysis special procedures

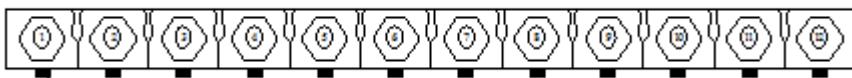


Fig. 2. The cross section of the measuring point

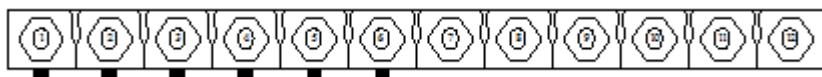


Fig. 3. L/4 section of the measuring point



Fig. 4. Network training performance curve

called Midas/Civil, and arrange the position of each parameter sensor settings. With bridge stress related sensor settings for example, according to the standard design load loading as calculating, according to the bridge, under the influence of live load and internal force of the envelope, and finally determined that the control section location, In the cross section of the measuring point in the arrangement of 12(see Fig. 2), in L / 4 we laid six experiments (see Fig. 3).

In order to Minimize the system spending and reduce the system load, we clean the sample data to test before we train the monitoring data, and in the light of the data are choosing different data normalization function, which makes great improvement convergence speed. With the speed of the wind load for example, we choose 1000 set of data to test, the consequence is as follows:

The train state function is described as Fig 5 and the regression function of the is plotted as Fig. 6.

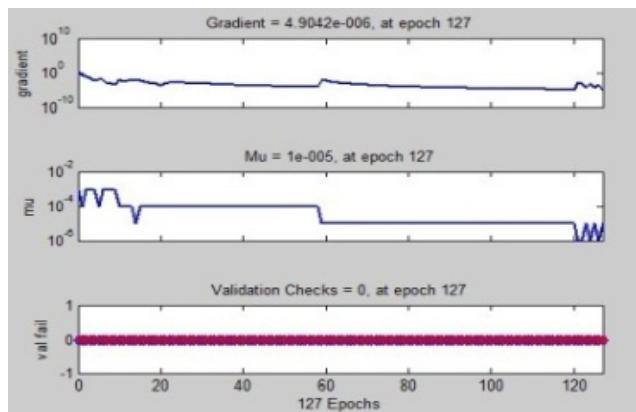


Fig. 5. Train state chart

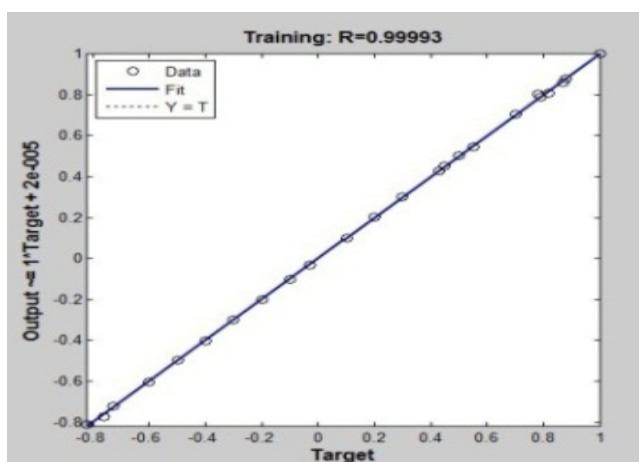


Fig. 6. Regression chart

We used 25 groups of the history monitoring data as a test data set

6 Summary

The method of monitoring of the bridge health state is based on two step neural networks ensemble which is not only good for the promotion of monitoring the bridges' health efficiency, but also the improvement of the veracities of the results. We can forecast the condition of the bridges by the intellective survey method at the real time, and we can also predict the life-circle of the bridges. The presentation of the two step neural networks ensemble makes the evaluation of the system under the complex conditions easier and more exact. The simulation verified the system can realize the safety assessment of the bridge, and help the intelligentization of the monitoring system with bridge health.

References

1. Hansen, L.K., Salamon, P.: Neural Network Ensembles. *IEEE Trans. Pattern Analysis and Machine Intelligence* 12(10), 993–1001 (1990)
2. Sollich, P., Krogh, A.: Learning With Ensembles: How Over-Fitting Can Be Useful. In: Touretzkyd, Mozer, Hasselmom (eds.) *Advances in Neural Information Processing System S* 8, pp. 190–196. MIT Press, Cambridge (1996)
3. Opitz, D., Maclin, R.: Popular Ensemble Methods: An Empirical Study. *Journal of Artificial Intelligence Research* 11, 169–198 (1999)
4. Ueda, N.: Optimal Linear Combination of Neural Network S For Improving Classification Performance. *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(2), 207–215 (2000)
5. Yang, G., Wu, X.-P., Song, Y.-X., Tian, S.-X.: Multi-Sensor Information Fusion Fault Diagnosis Method Based on Rough Set Theory. *Systems Engineering and Electronics*, 2013–2019 (August 2009)
6. Liu, J.-J., Xia, S.-P., Yu, W.-X.: An Ensemble of Neural Networks Based Incremental Learning Algorithm With Weights-Updated Voting. *Signal Processing*, 46–50 (January 2010)
7. Xie, J.-X., Cheng, C.-T., Qin, Y.: Multi-Step-Ahead Prediction Based on Parallel Neural Network Ensemble. *Computer Engineering and Applications* 29, 75–77 (2006)
8. Fang, Y.: Fusion Technology of Bp Neutral Network in Gas Forecast. *Coal Technology*, 107–109 (August 2011)
9. Qi, X.-Z., Liu, B.-J., Ji, H.-Y.: An Integrated Neural Network Ensemble Algorithm Based on Clustering Technology. *Computer Simulation*, 166–169 (January 2010)
10. Yu, S.-C., Chen, J., Ma, N.: A Method of Improving Neural Network Ensemble System S Generalization Error. *Microelectronics & Computer*, 105–107 (April 2009)

Classification of Rice Appearance Quality Based on LS-SVM Using Machine Vision

Xiai Chen¹, Shuang Ke², Ling Wang¹, Hong Xu¹, and Wenquan Chen¹

¹ School of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou, China

² Zhejiang Economic & Trade Polytechnic, Hangzhou, China

Abstract. A machine vision system was developed for rice quality detection in this paper. The main characteristics of rice appearance including area, perimeter, roughness and minimum enclosing rectangle were calculated by image analysis. The Least Squares Support Vector Machines was applied for the classification of head rice and broken rice. Genetic algorithm was used to optimize the parameters values of Least Squares Support Vector Machines. The robustness of this classification method was testified, and the experiment result shows that the head rice and broken rice can be effectively identified by Least Squares Support Vector Machines using machine vision.

Keywords: Rice, LS-SVM, GA, Machine vision, Appearance quality, Image processing.

1 Introduction

Commonly the quality of rice includes internal quality and appearance quality. The shape of rice is an important characteristic of rice appearance quality which directly affects the sales and prices of rice.

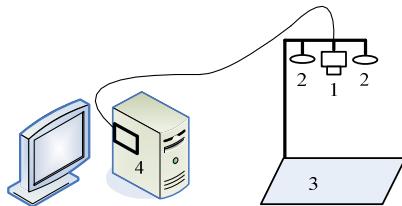
When the rice kernels are smaller than three-fourths of whole kernel they are defined as broken rice which seriously affecting the quality of rice, so it's very important to distinguish broken rice from head rice [1]. Currently, the appearance quality of rice is measured manually by inspector. This detection method is simple, but subjective and inefficient.

There is already a number of researching about rice quality detection [2,3,4]. However, most studies were about how to obtain the characteristic parameters of rice, while care less about the algorithm applied for classification of rice quality.

In this paper, machine vision was applied and the rice shape parameters were got after image processing and analysis. The classification and identification of rice grades was achieved by Least Squares Support Vector Machines (LS-SVM), and meanwhile the Genetic Algorithm (GA) was used to acquire the parameters of LS-SVM. The experiment result shows that the head rice and broken rice can be effectively identified by LS-SVM using machine vision.

2 Data Acquisition and Analysis

The hardware of Machine vision system applied in this experiment was shown in Fig. 1. It's mainly consisted of image acquisition, image processing and image analysis and display.



1. Carema 2. LED light source 3. Inspecting flat 4. Computer system

Fig. 1. Hardware schematic diagram of Machine vision system

After the RGB image acquired by frame grabber card, it was transform to gray image and then pre-processed by image processing algorithm including guassian filter, erosion, dilation, binarization, edge detection and filling. These algorithms were used to smooth and enhance image which prepared for image analysis.

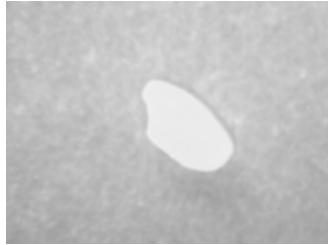
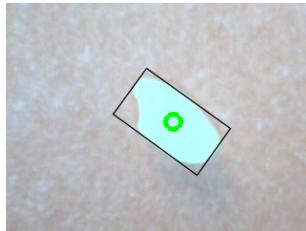
The eight characteristics of rice image were calculated by image analysis algorithms which shown in Table 1.

Table 1. Shape parameters of rice

ID	characteristic	Description
1	perimeter	perimeter of the rice image
2	Area	pixels number of the rice image
3	Roughness	$\text{Perimeter}^2/(4\pi \times \text{Area})$
4	MERL	length of minimum enclosing rectangle
5	MERW	width of minimum enclosing rectangle
6	MERA	$\text{MERL} \times \text{MERW}$
7	MERProportion	MERL/MERW
8	solidity	Area/MERA

Grayscale images of rice, binary image of rice after image pre-processing and the minimum bounding rectangle of rice calculated by image analysis are shown in Fig.2, Fig.3 and Fig.4.

In order to reduce the influence of background noise, image subtraction method was applied. First we got the last moment pure background image, and then subtract this background image from the rice image. The background image was updates periodically instead of fixed, which further improved the image detection and recognition accuracy.

**Fig. 2.** Grayscale image**Fig. 3.** Binary Image**Fig. 4.** Minimum enclosing rectangle

3 Classification Method

3.1 Least Squares Support Vector Machines

Least Squares Support Vector Machines (LS-SVM) is presented by Suyken s J. A. K [5]. The main differences of the SVM and the LS-SVM are as following, 1) The SVM solution is typically found by solving a quadratic programming problem where the LS-SVM solution can be found by solving a set of linear equations. 2) The SVM has inequality constraints where the LS-SVM has only equality constraints. 3) LS-SVM reduces the computational cost and easy to be implemented for on-line application [6].

Given a training data set of K points $\{x_k, y_k\}_{k=1}^N$, with input data $x_k \in R^n$ and output data $y_k \in R^m$, LS-SVM aims to build a classifier of the form as follows

$$y(x) = sign[\sum_{k=1}^N a_k y_k \psi(x, x_k) + b]. \quad (1)$$

Where a_k are positive real constants, b is a real constant and $\psi(x, x_k)$ is a Kernel function. The following optimization problem is formulated for classification in LS-SVM.

$$\min_{w,b,e} J_3(w, b, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2. \quad (2)$$

Subject to constraint as follows

$$y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, N. \quad (3)$$

To solve this quadratic programming, Lagrange multiplier is used as follows

$$L_3(w, b, e; a) = J_3(w, b, e) - \sum_{k=1}^N a_k \{y_k [w^T \varphi(x_k) + b] - 1 + e_k\}. \quad (4)$$

Where a_k are Lagrange multipliers, the conditions for optimality are given as follows

$$\begin{cases} \frac{\partial L_3}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N a_k y_k \varphi(x_k) \\ \frac{\partial L_3}{\partial b} = 0 \rightarrow \sum_{k=1}^N a_k y_k = 0 \\ \frac{\partial L_3}{\partial e_k} = 0 \rightarrow a_k = \gamma e_k \quad k = 1, \dots, N \\ \frac{\partial L_3}{\partial a_k} = 0 \rightarrow y_k [w^T \varphi(x_k) + b] - 1 + e_k = 0, \quad k = 1, \dots, N \end{cases}. \quad (5)$$

Eq. 5 can be written as the linear equations

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ e \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vec{1} \end{bmatrix}. \quad (6)$$

Where $Z = [\varphi(x_1)^T y_1; \dots; \varphi(x_N)^T y_N]$, $Y = [y_1; \dots; y_N]$, $\vec{1} = [1; \dots; 1]$, $e = [e_1; \dots; e_N]$, $a = [a_1; \dots; a_N]$.

With Mercer's condition

$$\Omega_{kl} = y_k y_l \varphi(x_k)^T \varphi(x_l) = y_k y_l \psi(x_k, x_l), \quad k = 1, \dots, N. \quad (7)$$

The equations related with a and b , Eq. 6 can be transformed into

$$\begin{bmatrix} 0 & -Y^T \\ Y & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix}. \quad (8)$$

Least squares method can be applied to calculate a and b from linear equations (8).

The Radial Basis Function (RBF) kernel $\psi(x_i, x_k) = \exp\{-\|x_i - x_k\|^2 / \delta\}$ was employed in the experiment, and the values of the kernel parameter δ and of the regularization parameter γ were determined by a grid search and cross validation to select those values that give the smallest error on the test data set.

3.2 Genetic Algorithm

Genetic algorithm is a stochastic global searching and optimizing algorithm which is used to solve complicated problems. GA including five steps of initialization, selection, recombination, mutation, and termination [7]. GA simulates natural evolution process, it start with an initial set of random solutions called population and use some measures of fitness to evaluate each chromosome. Through the recombination process, chromosomes with the higher probability are preserved. Then mutation was used to produce new variable values and prevent optimization process falling into local optimal. After several generations, the algorithms converge to the best chromosome, which hopefully represents the optimum [8].

Genetic algorithm was applied in calculating the kernel parameter δ and of the regularization parameter γ of LS-SVM in this paper. Compared to the grid search method, genetic algorithm greatly improves the problem solving speed.

3.3 Classification Result

The eight parameters calculated by image analysis of each 200 head rice and 100 broken rice were used as the training and testing data by cross validation. To test the robustness of this classification method, 5% and 10% noise was added respectively, which showed as Table 2.

Table 2. Classification result of rice detection

Algorithm	Recognition rate			Description
	Total	Head rice	Broken rice	
GA-LS-SVM	99.67%	100%	99%	Original data
GA-LS-SVM	99.33%	100%	98%	With 5% Noise
GA-LS-SVM	99%	99.5%	98%	with 10% Noise

Test results show that LS-SVM optimized by GA was very robust when used in classification of rice appearance quality.

4 Summary

This paper developed a machine vision system for rice appearance quality detection. After image acquiring, image pre-processing algorithm was employed to enhance the image quality. The characteristics of rice shape including area, perimeter, roughness

and minimum enclosing rectangle were calculated by image analysis. At last LS-SVM which optimized by GA was applied for the classification of head rice and broken rice, and then the robust of this method was testified. The result of this experiment show that broken rice can be distinguish from head rice effectively by LS-SVM, and the recognition accuracy of 99.67% was achieved.

Acknowledgment. This research work is supported by Scientific Research Fund of Zhejiang Provincial Education Department (Grant No. Y2011121959), Natural Science Foundation of Zhejiang province of China (Grant No. Y1100376).

References

1. Verma, B.: Image processing techniques for grading & classification of rice. In: 2010 International Conference on Computer and Communication Technology, pp. 220–223 (2010)
2. Yao, Q., Zhou, Y., Wang, J.: An automatic segmentation algorithm for touching rice grains images. In: 2010 International Conference on Audio Language and Image Processing, pp. 802–805 (2010)
3. Hobson, D.M., Carter, R.M., Yan, Y.: Characterisation and identification of rice grains through digital image analysis. In: Instrumentation and Measurement Technology Conference Proceedings 2007, pp. 1–5. IEEE Press, New York (2007)
4. Yao, M., Liu, M., Zheng, H.: Exterior quality inspection of rice based on computer vision. In: World Automation Congress 2010, pp. 369–374 (2010)
5. Suykens, J.A.K., Vandewalle, J.: Multiclass least squares support vector machines. In: International Joint Conference on Neural Networks, pp. 900–903 (1999)
6. Kuh, A.: Analysis of detectors for support vector machines and least square support vector machines. In: International Joint Conference on Neural Networks, pp. 1075–1079 (2002)
7. Goldberg, D.: Genetic algorithms in search, optimization, and machine learning. Pearson Publishing Group, USA (1989)
8. Guo, P., Wang, X., Han, Y.: The enhanced genetic algorithms for the optimization design. In: 3rd International Conference on Biomedical Engineering and Informatics, pp. 2990–2994 (2010)
9. Yao, M., Liu, M., Zheng, H.: Exterior quality inspection of rice based on computer vision. In: World Automation Congress (WAC), pp. 369–374 (2010)
10. Modi, C.K., Jain, K.R.: Quality Evaluation of Hydrothermal Treated Quicker Cooking Scented Rice by Quantification of Quickness of Cooking Time and Mechanical Strength Using Machine Vision. In: 2011 International Conference on Digital Computational Intelligence and Communication Networks (CICN), pp. 639–643 (2011)
11. Dai, X., Xiong, D.: Research on conjoint grains of rice based on machine vision. In: 2nd International Conference on Signal Processing Systems (ICSPS), pp. 760–763 (2010)

Research of University Disaster Recovery System Based on Virtualization Technology

Yue Li¹, Jian Sun², Zheng Zheng³, and Hong Shen^{1,*}

¹ Modern Technology and Education Center, Hebei United University,
Hebei Tangshan, 063000, P.R. China

² Office of Educational Administration, Hebei United University,
Hebei Tangshan, 063000, P.R. China

³ Information Technology, Jitang College of Hebei United University,
Hebei Tangshan, 063000, P.R. China

Abstract. With the continuous advance of the process of college information, the data center has become the hub of the school management, it will become more important. Currently how to ensure that the huge amounts of data in the data center to be a rational and efficient application has been a very important issue that Higher network management. This article aims to introduce the disaster recovery system that use of virtualization technology. It can quickly deploy the operating system and application software, achieve the Fast, efficient backup and recovery of university information system, improve the efficiency of data protection.

Keywords: Virtual technology, disaster recovery backup, Data center.

1 Introduction

With the maturity of the information application technology, various systems that the information technology brings to are very widely used in the daily work of universities and colleges. Such as academic system, network teaching system, office automation system, laboratory management system, E-mail system, etc. The data in these systems is the core of the system running. So, data security and reasonable application is critical. How to protect these data, protect the system to be stable and reliable operation, provide fast and reliable access to services for business systems are the most important of the college Information management system.

2 Virtualization Technology

2.1 Concept

Virtualization refers to the abstraction of resources is that put physical resources to the resources which can be manage in logic; it can break the barriers between the

* Corresponding author.

physical structure. All resources are running on a variety of physical platform transparently, resource management is running in a logical way. It can shield the difference of the hardware resources maximally by virtualization technology. Flexible allocation of the resources as needed, to reduce the overall cost.

2.2 The Advantages of Virtualization

The virtualization platform provides an automated control center to the administrator, it can significantly reduce the time and intensity of management staffs, greatly improved the efficiency. The old server use patterns is “one server, one application”, this mode cause the lack of resource utilization and serious waste of resources and facilities. At the same time, server administrator must cost a lot of time to maintenance and management.

3 Disaster Recovery Systems

3.1 Instruction

Disaster recovery system means that after the occurrence of variety of disaster, it still can Protect and enhance the normal application of computer information systems Maximize, can ensure the continuity of the business system, can achieve Transparent migration of the system. Thus, reduce any threat caused by system failures and other catastrophic events maximize.

3.2 Disaster Recovery System Based on Virtualization Technology

Virtual storage technology shield the differences of the storage hardware, integrate various devices into an integrated whole. Through the copy software of the virtual storage platform, to complete the data storage of the production center copy to the backup center equipment. It eliminates the difference between heterogeneous storage devices, to allow to a single resource composes of different storage systems to manage. Make full use of different storage equipments, to achieve disaster recovery capabilities of applications and important data.

The mode based on the virtual storage can ensure data integrity and consistency better, operation control simpler; eliminate the differences between the heterogeneous storage devices. Compared to the previous backup modes such as tape backup, database backup, imaging software backup, the mode based on the virtual storage have very significant advantages In the functions, real-time, security and other aspects.

4 Construction of Disaster Recovery System Based on Virtualization Technology

The last barrier of the data protection is backing them up. Fully integrated of storage, backup and disaster recovery technology, how to constitute a specific approach of integration of data disaster recovery system is as follows:

Through a proprietary backup network, it establishes the center of disaster recovery center in the local. For different applications development, different backup strategies and centralized backup achieve local data high-speed backup and recovery. Complete the backup progress effectively, Improve system manageability greatly, reduce the difficulty of management and the cost of administration system and improve the use and reliability of the backup information.

Deployment the same data backup disaster recovery center as system in different places, through technologies such as remote tape copy to achieve the backup of different application systems centralized backup. When in the abnormal situation, protect the data security maximizely; meet the reliability disaster recovery requirements fully.

To create a financial information service system, the following design should be fulfill.

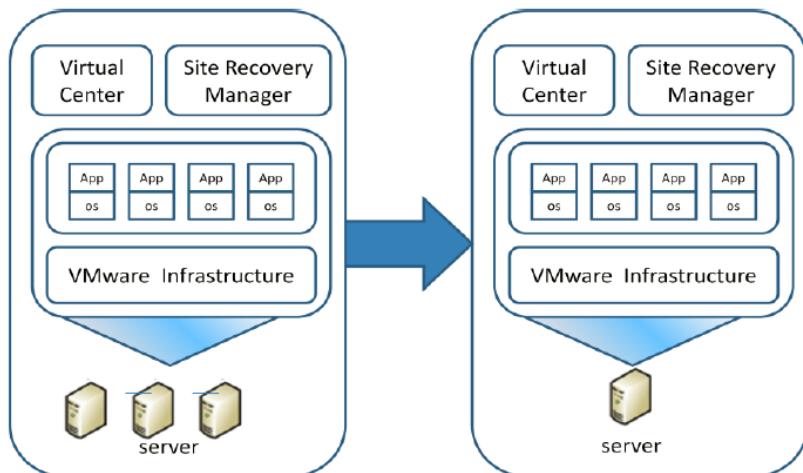


Fig. 1. The framework of virtualization disaster recovery

5 Application Strategy of Disaster Recovery Plan Based on Virtualization

University network managements can achieve the failover between the old data center and the disaster recovery site, through the disaster recovery that based on virtualization. Meanwhile, it also can manage the failover between the two active mutual recovery sites. The disaster recovery that based on virtualization can perform automatically and simplify the process of failover to a new data center, and then let failover of the data center plan becomes smoother

Based on the above characteristics, how to formulate a application strategy of disaster recovery plan based on virtualization will be a issue for the university network management and technical person needed to seriously consider. The

following is for the disaster recovery plan based on virtualization and the implementation plan to give a number of recommendations discourse.

5.1 Key Issues of Disaster Recovery Plan Based on Virtualization

5.1.1 Develop a Disaster Recovery Plan Based on Virtualization

The first step must be to develop a virtualized disaster recovery strategy; a great deal of convenience will be brought. It can reduce complexity and cost for managements, however decide when to implement the disaster recovery plan based on virtualization is a key factor and it is very difficult. So, assessment of the business about the university network is particularly important. Before implementation the plan, disaster recovery system maybe in a dispersed state, it is also difficult to manage. After the virtualization, the storage management can integrate multiple disaster recovery systems into a network environment. They can manage the entire network environment through a unified manner.

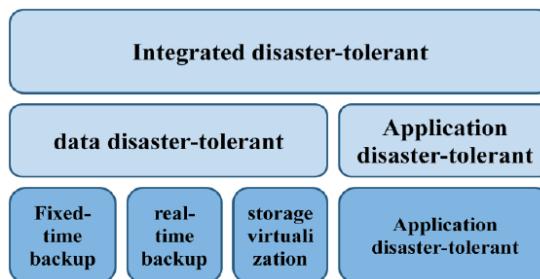


Fig. 2. Integrated disaster-tolerant

5.1.2 Reasonable Classification and Make Plan for Various Information Systems of the University Data Center

Various information systems of the university data center are very important and critical. According to their Critical level is more difficult to classify data. So, Reasonable classification will enables disaster recovery plan based on virtualization more efficient and more convenient and more efficient for management.

5.1.3 Understanding the Data Life Cycle

Virtualization Disaster Recovery has a very big advantage for the data life cycle in the optimization strategies. Network managers can unify management for all different information systems and any other system about them through the virtualization disaster recovery. It is facilitate administrators to achieve the migration of information systems.

5.1.4 Planning Virtualization Disaster Recovery Service Levels

Virtualization disaster recovery can balance the different levels of disaster recovery services, reduce the complexity of the different levels of disaster recovery and

improve the manageability of disaster recovery. The decide that the cost of virtualization for disaster recovery of the digital campus not a disaster recovery system's hardware, but a selection that virtualization disaster recovery and solutions. The virtualization disaster recovery service-level planning determine the decision of a virtualized disaster recovery platform and program selection.

5.1.5 Consider the Security of Virtualization for Disaster Recovery

Virtualization disaster recovery system can supply intensification and secure computing environment, it can ensure the confidentiality of sensitive data. Security is the implementation of virtualization; disaster recovery plan needs to focus on consideration. So, we should pay more carefully considered to the virtualized disaster recovery environment.

5.2 The Main Points of the Implementation of the Recovery Plan of Virtualization Disaster

To develop and implement the virtualization disaster recovery plan needs to focus on the following combination of research the disaster recovery of a university data center and the actual situation for a university data center how to carry out the operation of disaster recovery.

The recovery system of virtualization disaster need a better compatibility than Server and Client, and it must be able to support multi-operating system, backup and resume various popular database system and all types of storage device. This system has wide compatibility to meet the demands of the entire customers.

To operate the recovery system of virtualization disaster must have the simplest methods of operation. Backup the data resources through a unified graphical interface. And uniform definition of backup strategies and recovery strategy. And then definite uniformly of backup strategies and recovery strategy for reducing the complexity of the whole system to adapt to different needs.

The function of virtualization disaster recovery system must be perfected, it is must carry out a complete data backup off-site disaster recovery and data recovery. The database systems and file systems are achieved online backup.

6 Attention to the Disaster Recovery System Development and Construction

6.1 Strengthen Enterprise Security Management

Disaster recovery system can only prevent disaster in a certain extent, but in the event of a disaster in human factors and environmental factors cannot be neglected. Due to poor management, the improper operation of staff and the room environment unexpected accidents caused by system failures caused often. In terms of guard system disaster, A well-functioning security system is essential, strengthen the

management of personnel, network Security, establishment of a secure system is the most important for ensure the data security.

6.2 Local Conditions the Construction of Systems

Need to build their own disaster recovery centers according to their actual demand for body. Maximize the use of advanced technology, According to their needs has become a basic criterion for building a disaster recovery program. Disaster Recovery Center can be divided mainly prepared by the center and two-center two modes. Two-center load sharing, mutual disaster, to take local conditions the principle, to develop system construction program of specific disaster recovery.

6.3 The Construction of a Disaster Long-Term

Disaster recovery of data for disaster protection, in recent years, with the rapid development of University information, university information system of relevant information and data grow exponentially. Because of the disaster recovery center assume the user the most critical core business and it is no doubt that their importance. All of this determine the disaster as a systems engineering, can't able to be done overnight. With the development of disaster recovery technology, future disaster recovery center will be more quickly and easily, data and business will be more secure.

7 Summary

In summary, with the continuously development of the university information system and the scale of the further expansion, put forward higher requirements on the university information system and the construction of the data center. Prevent the occurrence of sudden accidents, through building disaster recovery system based on virtualization technology. Provide a solid foundation for the rapid development of university information.

References

1. Zhou, H.: Research and construction of disaster recovery system based on virtualization technology. Zhejiang Electric Power, 54–56 (August 2011)
2. Dong, Y.: Environment construction of experimental teaching center based on virtualization technology. Experimental Technology and Management 28(3), 299–302 (2011)
3. Wang, S., Liu, B.: Virtualization disaster recovery scheme for data center. Computer Security, 53–54 (July 2010)
4. Liu, Q., Zheng, W., Chen, K.: Research on disaster recovery system based on migration of virtual machine. Journal of Chinese Computer System 31(10), 1994–1998 (2010)

5. Kang, X., Yang, Y., Du, X.: Application of virtual storage technology in disaster tolerant system. *Computer Engineering* 35(21), 36–41 (2009)
6. Liang, H.: Building disaster recovery backup system and ensuring the enterprise's data security. *Enterprise Science and Technology & Development* 16(286), 94–95 (2010)
7. Zhou, B., Zhang, Y.: Analysis and comparison of virtualization technologies performance. *Modern Computer*, 126–129 (May 2009)
8. Wen, J., Gong, D.: Application research on virtualization of DRP for telecommunication. China Academic Journal Electronic Publishing House, 16–20 (August 2009)
9. Yang, Y., Yao, W., Chen, Z.: Review of Disaster Backup and Recovery Technology of Information System. *Journal of Beijing University of Posts and Telecommunications* 33(2), 1–6 (2010)
10. Qian, L., Li, H., Xie, X., Chen, Z.: The application research of virtualization techniques for high performance computer systems. *Computer Engineering & Science* 31(A1), 307–311 (2009)

Research of Self-study English Pronunciation on WindowsCE

Shen Hong¹, Li Yue¹, and Wang Jinfang^{2,*}

¹ Modern Technology and Education Centre, Hebei United University,
Hebei Tangshan, 063009, P.R. China

² College of Foreign Languages, Hebei United University,
Hebei Tangshan, 063009, P.R. China

Abstract. The paper elaborates the design procedure of the software, Self-study English Pronunciation, which is installed on Pocket PC of ARM series. Its database design includes the design of database table, memory space, storage means and data flow. The frame design includes the design of MFC framework, multiple view and packet package. Through running the software, the user can implement self-study according to the chapters and sections of syllabus, choosing any units from the contents such as text, sentence and vocabulary, and recording or replaying their own speech sounds to learn.

Keywords: English Self-study, platform of WindowsCE, Software Design.

1 Introduction

Speech sound is the most directive, convenient and natural communication way. Among the media of computer and human, the input of pronunciation is difficult to develop, and speech recognition is the nucleus technology of computer intelligence, by which computer makes pronunciation code into correspond document or command. The technology of speech recognition has been gradually developing in recent forty years, and plays more and more important role in daily life. The technology has primarily been used in such fields such as dictation machine, speech sound data search, speech control exchanger, oral translation, oral dial, internet browse, etc.

The self-study software of English pronunciation is independent. The user can choose any chapters from the contents, copy and play his own speech sound of text, sentence, and vocabulary repeatedly. The test in each chapter can be used as self estimation. The user can record the study result and acquisition on the notebook, check the notebook in any time, and switch to the last study interface to continue learning.

2 Demand of English Self-study Pronunciation

The self-study software of English pronunciation should have the following functions:

* Corresponding author.

2.1 Learn and Practise Based on Chapters

When the user enters the main menu, and chooses the parts (volume, chapter and unit are arranged according to syllabus) he likes, the six parts, text, sentence, vocabulary, unit test, dialogue and sentence reorganization, are available.

2.2 Interactive Translation of English and Chinese

The user can search Chinese and English explanations of any words from the database. It is available to search the Chinese interpretation of English word, and the interface will change to the page that the English word is in and turn to 'Choose to Learn' to practice.

2.3 Self-defined Sentence Practice

The user inputs some sentences and stores them in the database DIYSentenceTable. Click the begin button to practice, the system will call out the sentence text from DIYSentenceTable and show them on the screen. The user can add, delete, check and revise these sentences, record the pronunciation of these sentences repeatedly and view the estimation score of pronunciation.

2.4 Review Learning Record

Users can browse their learning record, modify the remark of their blogs, delete one or all of their record, skip to the learning interface of one of records and continue to learn the unfinished.

2.5 Guide to Learn

Users can search any topics if they need help, and know all the self-introductions of main characters in the text.

3 Database Design

3.1 Table Design

According to the demand of the software, there are ten kinds of results that users ask to feedback: the contents of learning blog, the test of chapter, the sentence reorganization, the text, the dialogue, the Chinese and English interpretation of vocabulary, all of which are used to determine the number of forms. Besides, how to put various coursewares into the database and arrange an effective and brief database is one of essential purposes for an embedded system.

The table design and the relations of tables are restricted by the teaching materials. That is to say, the database must embody the characteristics of the teaching materials structure that has a clear structure of volume, chapter, and unit.

3.2 Compress Store Space

The courseware needs a large amount of multimedia files, such as text film, large text, audio files, etc. These materials can be highly compressed and searched flexibly in databases, such as MS SQLServer and so on. But when it is WinCE 3.0, generally speaking which has 4G space, it is unwise to put the courseware in the database. The user usually downloads files with portable equipment by connecting desk PC, which is more common than moving equipment and it is also economic and fast. Therefore all large data files should put into their own folders. The database gives the relative route name. The user can download any chapters' folders and install relative route. When the system runs, it will take out the route name cluster. And put the installation route name before the route name cluster, the whole file route will be got.

3.3 Measurement of Storage Way

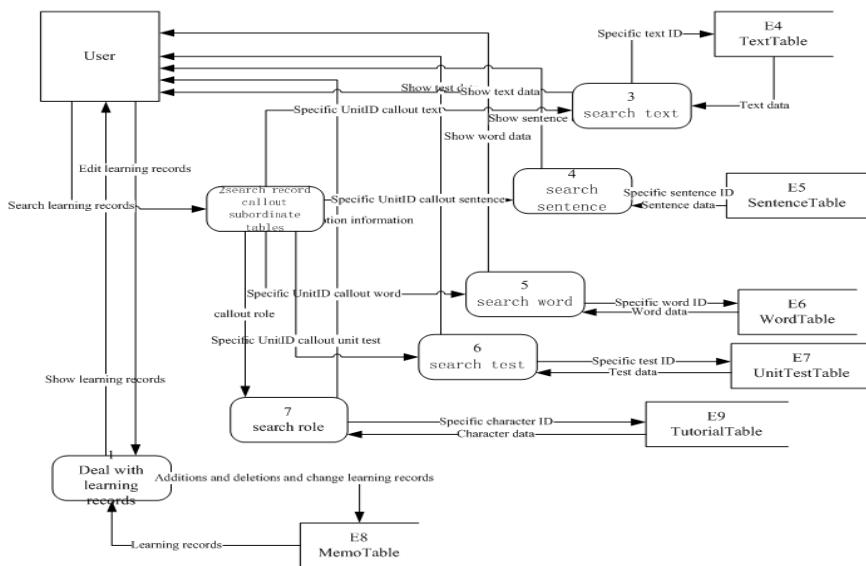
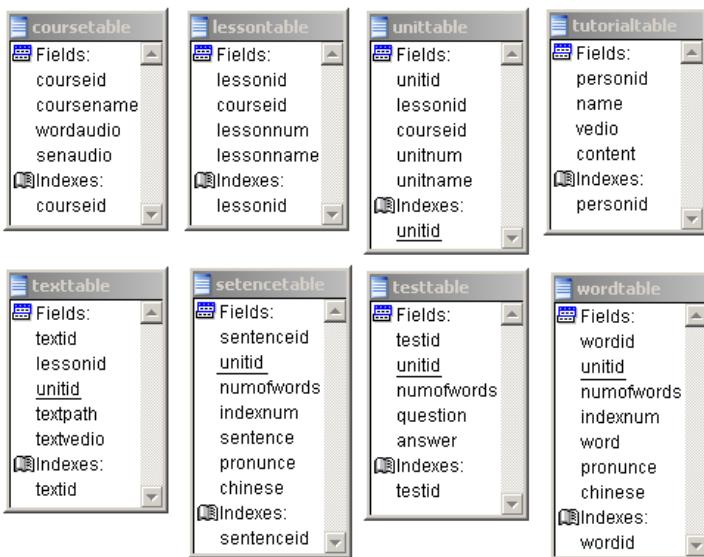
The contents of teaching materials are the way of only read, which is good for compiling and tiding databases on PC. Only two forms does the user add, delete, check and change learning blog table and self-making sentence table. WinCE offers special CE database access function interface, which is easy to operate. But the CE database is only identified by the machine itself, which has no way to switch and can not be seen in browser. The .dbf form file, which is simple and like the style of text, can be opened and edited by MSFoxpro. Self-defining a database of only read engine and search limited several types of byte. The criteria of program supporting .dbf file have two versions of Foxpro9 and dBase5.

3.4 Data Flow Diagram

The graph above shows the first grade flow diagram (including 0, 1 and 2 three grades). It also shows that user calls out learning materials by invoking different database tables and ultimately returns different results. In process 2 (search record callout subordinate tables), it chooses learning contents and marks specific unit with UnitID, callouts text materials and sends them to the following progress 3 (search text). Process 3 takes out the text materials of UnitID from TextTable, and turns them into the materials of interface, and ultimately goes back to show them to users.

3.5 Foxpro Design (Only Read)

The following chart is the eight database tables of Foxpro implementation.

**Fig. 1.** Data flow diagram**Fig. 2.** Eight database tables

4 Database Implementation

There are three database programming in the operating system Windows CE 3.0.

- (1)VB+ACCESS CE programming
- (2)DAOCE connection(above VB+SQLServerCE1.1)
- (3)And specific invoking function of WinCE database table.

The first two ways are offered to Visual Basic to program in the format of control. It is known that VB control can lead in VC project. But Evc3.0 strictly forbid the use of OLE DB. The third database programming actually is not the general database function because its progressing objects are the specific database tables on WinCE, which are not opened or compiled on PC.

A large amount of database information in the software of pronunciation learning should be downloaded from PC (PC gets resources from internet). That is to say, all of the databases should be compiled on PC, and they only are designed to access and have no interface to modify. By this, the difficulty of compiling databases is lower. The reason why Foxpro (table files are marked by .dbf and database project files by .dct) is chosen mainly lies in the simplicity of the .dbf file. Basic access files form into a database engine pack (see DBEngine.h and DBEngine.cpp), which does not need the support of document index and can be called directly from interface. Their attributes are private, have less codes and high effective access efficiency. Foxpro is not the database of relationship type, does not support integration primary key, which bring a little difficult to the building of database resources and the checking of database integrity. But Oracle has ideal products aiming to the third party, Oracle9iLite, 750K in size (PC development version 340MB), which is convenient to realize the data synchronization between roaming terminal and Oracle9i Server and support complex relationship database.

5 Frame of Self-study Pronunciation Software

5.1 MFC Frame and Multiple Views

PocketPC platform interface is limited and has no minimum icon. In order to realize the five function blocks, the switch of multiple views is difficult to handle.

MFC framework preserves the merits of Win32 API message image, but is not suitable to correspond multiple views with single text (eVC does not support MDI). MFC is chosen as the frame and CDocument container in MFC is not included.

In the engineering of MFC, the program entrance is CWinAPP case named CALLEnglishAPP, which produces MainFrame case and CALLEnglishDoc case. Under MainFrame there are three juxtaposition pages: MainView, MemoView and TutorialView. CALLEnglishAPP structures dynamic interface to cater for the demand of users. User can choose any learning interfaces in the main interface. The estimation marks are shown by a dialog box ShowScoreDialog if the user wants to know. In the process of learning, the user can switch to MemoEditdialog and save learning log.

Under MainFrame four first views are formed, which switch by SwitchView function control. And at the same time, context parameter maintains the context information of each views, realizing the correspondence between multiple views and single document.

Under the first grade views, the pop dialog box CDialog is used to realize the complex view invocation. In the dialog box, the database engine is shown by cases. When the dialog box is shut down, the internal storage that the cases occupy should release immediately.

Interface calls out self-definition package or class to complete single specific function. Some packages or classes need invoke pronunciation algorithm offered by lab. The algorithm is carried in operational procedure by .dll dynamic interlink to reduce the overhead of system internal storage.

In the software, the programs built on MFC frame have less hierarchy. The interface class in the picture should response to the user information and deal with it so that they are control classes too. The interface calss should form case database engine object, using the ELT operation to visit specific table, so that it has some characteristics of entity class.

5.2 Packaging

The three encapsulation packages are used to call out .dbf database file, .dat lab self-definition pronunciation integration file, and record or play pronunciation.

(1)DBEngine Package

The self-definition database engine package DBEngine has three object classes and a structure entity. DBTable is used to invoke .dbf database file and save table field that file describes in the structure entity DBField, and search file to get the record number content and save the record number content in the byte block RecordSet. DBRecord can operate the byte block, such as take out a record, or some of the bytes of the record, etc. DBProp is aim to operate the concrete byte information. For example, take out of the value from it.

(2) DatFile Package

The package DatFile is used to invoke the courseware pronunciation materials on physics storage. CDatfile is responsible to read out the sentence pronunciation data that is in the .dat pronunciation integration file and in the selected order.

(3) WaveThread package

The package has two user interface thread classes which are relevant to waveform audio equipment. ThreadPlay is to play pronunciation and ThreadRecord to record.

6 Realization Tool of Self-study Pronunciation Software

WindowsCE.NET

WinCE.NET is an important measure for Microsoft Company, who push forward .NET strategy into the domain of imbedding.

(1) The minimal operation system volume: abundant function is modularized in granularity. The smallest configuration of WinCE .NET is only 200KB.

(2)Board Support Package(BSP): Windows CE includes various BSP, support extensive CUP choices and their emulation technique: ARM, MIPS, SHx, X86, which reduce the time that operational system functions on hardware.

(3) Open communication platform: support LAN, Ethernet (802.3), IrDA, WAN, WLAN, Token Ring(802.5), UPP(for example USB Host) etc. , support true time mass media communication and ftp, telnet, web service and so on. The internet safe supports SSPI,SSL and Kerberos, etc.

(4) Abundant multimedia supports: support the latest multimedia experiences, including Microsoft DirectX 8 API, Windows Media 8 Codec and ActiveX.

(5) Tool for developer: Platform Builder is an IDE, including a SDK derived tool, supports EVC4.O and Visual Studio.NET, and offers an integrity development environment for the Web service and development of Microsoft.NET Compact Framework.

(6) Page environment virtual internal storage management: application program need not worry about the real application position of internal storage. The operating process occupies 64KB heap space, which has added another 32KB page environment and has double space than before.

7 Summary

It is a gradual process that multimedia such as sound; picture and film are played and shown on WinCE platform, expiring the short term skip from bottom layer device interface to ActiveX. The article briefly describes a integrality module packaging and bottom database design of an English studying software WinCE3.0 version, hoping the software on small portable equipment play important role and become an beneficial supplementary of college English teaching.

References

1. Microsoft Windows CE Programmer's Guide. Microsoft Press, America (1998)
2. Hamada, Miki, Satoshi, et al.: Automatic evaluation of English pronunciation based on speech recognition techniques. In: EUROSPEECH 1989, pp. 1421–1424 (1989)
3. Huggins-Daines, Kumar, Chan, et al.: Pocketsphinx: A Free, Real-Time Continuous Speech Recognition System for Hand-Held Devices. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2006 (2006)
4. Dollimore, J., Kindberg, T., Coulouri, G.: Distributed Systems: Concepts and Design, 4th edn. Addison Wesley, English (2005)
5. Xu, P.X.: Voice dialing system, user interface design. Technology Trend, p. 232 (September 2010)
6. Liu, J.B., Yuan, H., Mao, Y.Y.: Based on mobile learning self - explore - Collaborative Learning Mode. China Modern Educational Equipment, 141–143 (December 2007)
7. Zuo, G.H.: Research of Streaming Media Based on Windows CE and Key Technologies. Tianjin University (2006)

8. Wang, P., Liu, J.Y., Chen, Q.P.: Analyzer based on the color WinCE research and design. *Electronic Measurement Technology*, 49–51 (May 2011)
9. Song, F.F., Song, X.L., Ma, Y.Q.: The Research of the Scoring Mechanism of Spoken English Learning System Based on the Technology of Speech Recognition. *Computer Knowledge and Technology* 5(7), 1726–1728 (2009)
10. Qiao, B., Wu, Q.L., Yin, Y.M.: Implementation of Speech Recognition Algorithm in VC++. *OME Information* 28(4), 50–54 (2011)
11. Huang, X.Y., Yu, W.P.: The Application of Speech Recognition Technology in Foreign Spoken Language Learning. *Computer Systems Applications*, 18–21 (June 2006)
12. Zhang, X.Y., Wang, H.J., Liang, J.H.: LCD system based on embedded processor and Windows CE system. *Electronic Measurement Technology* 31(7), 76–80 (2008)

Study of Linearization for Spectral Imaging Device

Songhua He

Institute of Media and Communication Shenzhen Polytechnic, Shenzhen, 518055, China
hdh1818@szpt.edu.cn

Abstract. To improve spectral reconstructive accuracy for spectral imaging system based on trichromatic digital camera, a new radiation intensity-based linearization technique was put forward. The accuracy of the method was verified in experiment, whose performance was similar with spectral sensitivities-based technique, but better than Luminance-based technique.

Keywords: Linearization, spectral imaging, trichromatic digital camera.

1 Introduction

The price of Special Multi-spectral imaging device is expensive, and working conditions are complicated, so some of research began to capture multi-spectral images using a conventional trichromatic digital camera combined with either absorption filters or multi-illumination [1,2]. But the CCD in consumer digital camera is not scientific CCD comparing with Special Multi-spectral imaging device, which are disturbed by noise easily and the information processed by CCD need complicated transmit and compress and then transport to output device, actual digital camera responses are non linearity. Frequently-used spectral reconstruction algorithms are built based on linear mode of imaging system, for example, pseudoinverse, winner estimation, PCA, and so on [3]. If we use directly actual digital responses to reconstruct spectrum, the performance of spectral reconstruction and chromatic accuracy is low and can't satisfy need of multi-spectral image. This article set up a new linearization method of spectral imaging device based on Thomson algorithm [4,5] and verify it in experiment.

2 Camera Response Models and Multi-spectral Image Capture

During the last decade, digital imaging systems such as digital cameras have been widely used to record scenes. A simple trichromatic camera response model can be expressed as:

$$C_i(x, y) = \int_{400nm}^{700nm} E(\lambda)S_i(\lambda)P(x, y, \lambda)d\lambda \quad (1)$$

where C_i is the camera responses, $E(\lambda)$ is the spectral radiant distribution of illuminant, $P(\lambda)$ is the spectral reflectance of an object pixel, and $S_i(\lambda)$ is the spectral sensitivity of the camera's i th channel.

Equation 1 is ideal camera response model. Furthermore, various noise should be considered which usually includes dark current noise, shot noise, read noise and quantisation noise.

$$C_i(x, y) = \int_{400nm}^{700nm} E(\lambda)S_i(\lambda)P(x, y, \lambda)d\lambda + n_i(x, y) \quad (2)$$

Where $n_i(x, y)$ is noise of camera's i th channel.

A non-linear relationship between camera input and output responses has often been introduced by the manufacturing or electronic process. In the non-linearity model, the camera response for a pixel of the i th sensor type pixel is given by:

$$U = \Psi(C) = \Psi\left(\int_{400nm}^{700nm} E(\lambda)S_i(\lambda)P(x, y, \lambda)d\lambda + n_i(x, y)\right) \quad (3)$$

where Ψ is a monotonically increasing non-linear function, U is actual camera responses, this is R,G,B values read from photographic image.

Traditional trichromatic imaging systems capture and represent information about real scenes using three sensors which respond broadly in the red, green and blue regions of the visible spectrum. Traditional trichromatic images (or sometimes termed RGB images) are both device and illuminant dependent. Although, These problems can be solved by applying a process called device characterization which is defined as finding the relationship between the device color space and the CIE system of color measurement. But XYZ color space is metamerism, when two color samples have the same tristimulus values for a specific combination of illuminant and observer, but the reflectance curves are different [6].

It is well known that the only way to assure a color match for all observers across changes in illumination is to achieve a spectral match. Imai and Berns put forward a new spectral structure system in which a conventional trichromatic digital camera combined with either absorption filters or different light sources [7]. Multi-filter or multi-illuminant approach can make the image acquisition easier than the traditional monochrome camera and interference-filter-based multi-spectral acquisition because it can provide three channels per imaging shot. This trichromatic camera approach also has relatively low cost since the cost-performance relation of commercial trichromatic digital cameras has decreased rapidly.

Spectral reconstruction algorithms used in traditional multi-spectral image capture, are also used in spectral structure based trichromatic digital cameras. the spectral reflectance of each pixel can be estimated using a priori spectral analysis with direct measurement and imaging of color patches to establish a relationship between the digital counts and spectral reflectance.

A set of spectral reflectances, r , is measured and then the corresponding set of eigenvectors, e , is calculated by principal component analysis. Then, the set of

coefficients, W , corresponding to the eigenvectors, e , is calculated using the spectral reflectances, r . A relationship between coefficients and digital counts, C obtained using trichromatic camera and absorption filters can be established by the equation

$$A = WC^T (CC^T)^{-1} \quad (4)$$

where T denotes transpose matrix. The matrix A can be used to calculate the coefficients, W , from digital counts to reconstruct the spectral reflectance.

3 Linearization of Spectral Imaging Device

As analysed in equation 3, there is a non-linear relationship between camera input and output responses. Because most of spectral reconstruction algorithms are built based on linear relationship, if we use actual digital response values to reconstruct spectrum in linear relationship, it will produce big errors.

Consider that a spatially uniform surface of known spectral reflectance $P(\lambda)$ is captured under an illuminant with known spectral power distribution $E(\lambda)$ by a three-channel imaging system with spectral sensitivities $S_r(\lambda)$, $S_g(\lambda)$ and $S_b(\lambda)$. Thus, if we represent the variables by discrete samples at uniform intervals of wavelength λ the raw channel responses R' , G' and B' for the red, green, and blue channels respectively are given by the following:

$$\begin{aligned} R' &= \sum E(\lambda)S_r(\lambda)P(\lambda) \\ G' &= \sum E(\lambda)S_g(\lambda)P(\lambda) \\ B' &= \sum E(\lambda)S_b(\lambda)P(\lambda) \end{aligned} \quad (5)$$

We assume that the channels are subject to a nonlinearity Ψ to generate the actual output responses R , G , and B . Eq(6) shows an example for the blue channel; similar nonlinearities are assumed to exist for red and green channels as follows:

$$B = \Psi(B') \quad (6)$$

Actually, we know actual output responses R , G and B read from images, and aim to look for raw channel responses which are linearization in Eq(5), so eq(6) is expressed as follows:

$$B' = \Gamma(B) \quad (7)$$

Where Γ is a monotonically increasing non-linear function, the aim of linearization is to build Γ function. However, that to compute the raw channel input B' , the spectral sensitivity of the channel is required. Actually, in most cases, we don't know spectral sensitivity of the channel, besides methods based on a knowledge (or estimation) of the spectral sensitivities of the channels. Methods based on the luminance were put

forward[8]. luminance Y is often used instead of the raw channel responses B' to determine the nonlinearity function Γ .

$$Y = \Gamma(B) \quad (8)$$

Where $Y = \sum E(\lambda)V(\lambda)P(\lambda)$, $V(\lambda)$ is the luminous efficiency function of each surface.

Here, we put forward a new linearization based the radiation intensity of a set of neutral samples based on radiation intensity is proportion to raw digital camera responses[9].

Radiation intensity I is often used instead of the raw channel responses B' to determine the nonlinearity function Γ .

$$I = \Gamma(B) \quad (9)$$

Where $I = \sum E(\lambda)P(\lambda)$.The nonlinearity can be modeled using a power law function or polynomial regression with least-squares fit as follows:

$$B' = B^r \quad (10)$$

$$B' = aB_i^n + bB_i^{n-1} + \dots + cx_i + d \quad (11)$$

4 Experiments

In our experiments, we used a Nikon D700 digital camera ($4,256 \times 2,832$ pixels, built-in R, G, B array sensors, 8bits per channel). The spectral radiant power of the illuminant used in this imaging system were measured. We used data measured by Christian Mauer as The spectral sensitivities of each of the color channels in digital camera system. Two imaging targets, 140 GretagMacbeth ColorChecker SG (abbreviated in CCSG) and 24 ColorChecker RC(abbreviated in CCRC) were imaged. 14 gray colors located in the center of CCSG were used for Linearization. CCRC as training samples and testing samples were verified for accuracy of spectral reconstruction. The spectral reflectance factors of the patches on the two charts were measured using an GretagMacbeth Spectrolino spectrophotometer. During the experiment the camera and lighting positions were fixed, and the RGB values of the CCSG achromatic samples and CCDC were measured with each in the center of the camera's field of view.

In our experiments, we used three different linearization methods. Here, we only give the steps of radiation intensity-based technique which are as below:

The radiation intensity I were computed for 14 CCDC gray samples under the actual light source that was used to illuminate the samples during imaging, and normalized in the range 0–1 based on the largest radiation intensity.

The actual camera responses of 14 CCDC gray samples were measured and normalized in the range 0–1 based on (255,255,255).

For each camera channel, a n-order polynomial or γ power law function was used to fit the relationship between the actual camera responses and the radiation intensity for the achromatic patches, then function Γ was built.

Apply the estimates of function Γ to the measured camera responses for the CCDC samples to yield the linearized RGB values for those values.

Perform spectral reconstruction for CCDC samples based on PCA.

Thus for each technique we attempt to obtain linearized camera responses and then use PCA to reconstruct the spectral reflective factor. All computations were performed using MATLAB, RMS errors and CIELAB ΔE^*ab were taken into account to test the performance of spectrum and color differences between original and reconstructed spectra.

5 Results

Fig. 1 show compare between non-linearity and linearization based on Radiation intensity. The plot on the left-hand side of Fig. 1 show the normalized R values for the gray samples plotted against the respective normalized measured radiation intensity. The solid line in each panel shows the polynomial fit to the data. it is evident that the linearity of NIKON D700 is low. However, these polynomial functions were used to subsequently convert all actual camera responses to linearized camera responses before further processing. The plot on the right-hand side of Fig. 1 show the normalized linearized R values for the achromatic samples plotted against the normalized measured radiation intensity for those samples.

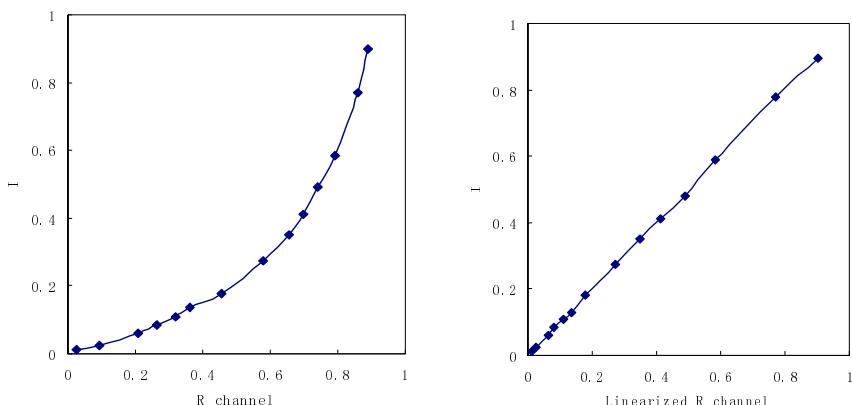


Fig. 1. Radiation intensity I of 14 achromatic patches plotted against the spatially averaged camera channel RGB responses (blocks) for each patch and fifth-order polynomial fit (solid lines) (left column) and the linearized camera responses (blocks) converted from the polynomial fit (right column).

Table 1. Estimation of the exponent using each of the three different linearization techniques

Linearization method	Estimate of exponent		
	Red channel	Green channel	Blue channel
Spectral-sensitivities-based	2.193	2.321	1.962
Luminance- based	2.176	2.429	1.97
Radiationintensity-based	2.187	2.34	1.98

Table 1 shows the values of the exponent that were estimated using each of the three linearization techniques. We found the exponent of each channel are similar for three linearization, these mean power law functions are similar, there are similar relationship between different raw digital camera response and actual digital camera response. Just as table 2 shows spectral reconstruction performance for CCDC samples based on linearization by three various techniques. Because the PCA spectral reconstruction method is not perfect, color differences will still be obtained even if the input/output nonlinearity is estimated perfectly. After linearized, Spectral reconstruction performance either spectral RMS or CIELAB ΔE are better than non linearity. Generally, spectral-sensitivities-based linearization and Radiation intensity-based linearization are quite similar in spectral and chromatic accuracy, which are better than performance of Luminance- based linearization.

Table 3 and table 4 are different with table 2 in function fit. Table 3 use 3rd polynomial fit, table 2 use 5th polynomial fit. It's obvious that the level of polynomial fit is higher, the accuracy of linearization function is higher. The accuracy of linearization techniques using a power law function is between 3rd and 5th polynomial fit, higher than 3rd, lower than 5th .But a power law form of the nonlinearity is simpler than polynomial fit, it only needs estimate a exponent comparing with mutli-coefficients of polynomial fit. There are same conclusion that spectral-sensitivities-based linearization and Radiation intensity-based linearization are better than Luminance-based linearization. This means building linearization by Radiation intensity instead of raw digital camera response is feasible, because in most conditions we don't know spectral sensitivities of camera.

Table 2. Comparison of spectral reconstruction performance with different linearization techniques using a power law function

Linearization method	Maximum ΔE	Minimum ΔE	Median ΔE	Maximum RMS	Minimum RMS	Median RMS
Non linearity	29.47	5.74	12.5	0.3196	0.0223	0.0803
Spectral-sensitivities-based	16.9	1.76	6.70	0.1280	0.0175	0.0515
Luminance-based	17.43	2.23	6.93	0.1267	0.0181	0.0514
Radiation intensity-based	17.01	1.74	6.72	0.1272	0.0176	0.0514

Table 3. Comparison of spectral reconstruction performance with three different linearization techniques using 3rd polynomial fit

Linearization method	Maximum ΔE	Minimum ΔE	Median ΔE	Maximum RMS	Minimum RMS	Median RMS
Non linearity	29.47	5.74	12.5	0.3196	0.0223	0.0803
Spectral-sensitivities-based	19.07	1.3	7.29	0.1601	0.0049	0.0509
Luminance-based	19.04	1.52	7.31	0.1611	0.0051	0.0511
Radiation intensity-based	18.89	1.54	7.22	0.1604	0.0050	0.0509

Table 4. Comparison of spectral reconstruction performance with different linearization techniques using 5th polynomial fit

Linearization method	Maximum ΔE	Minimum ΔE	Median ΔE	Maximum RMS	Minimum RMS	Median RMS
Non linearity	29.47	5.74	12.5	0.3196	0.0223	0.0803
Spectral-sensitivities-based	17.98	1.34	5.96	0.1726	0.0077	0.0521
Luminance-based	17.86	1.23	5.96	0.1730	0.0075	0.0520
Radiation intensity-based	17.82	1.21	5.90	0.1716	0.0076	0.0518

6 Conclusion

This study investigated techniques for accounting for the nonlinearity of the input-output response of a digital camera system. A new radiation intensity-based linearization technique was put forward. Three various Linearization was carried out for an Nikon D700 camera and the Macbeth ColorChecker RC achromatic samples, and the performance of spectral reconstruction was tested using the ColorChecker DC samples. The results showed the accuracy of intensity-based linearization technique are similar with Spectral sensitivities-based technique, but better than Luminance-based technique. In our experiment, the test of linearization technique is only for three imaging channel, so spectral reconstruction and chromatic accuracy is lower than multi-spectral imaging system. if using a conventional trichromatic digital camera combined with either absorption filters or multi-illumination, the spectral reconstruction results will improve highly accuracy.

References

1. Imai, F.H., Berns, R.S.: Spectral Estimation Using Trichromatic Digital Cameras. In: International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives, Chiba, Japan, pp. 42–49 (1999)
2. Chen, Y., Xu, H., Zhang, X., Ronnier, M.: LuoStudy of Spectral Reconstruction Based on Digital Camera. *Acta Optica Sinica* 29(5), 1417–1419 (2009)
3. Ribés, A., Schmitt, F., Pillay, R., Lahanier, C.: Calibration and spectral reconstruction for crisatel: An art painting multispectral acquisition system. *J. Imag. Sci. Tech.* 49, 563–573 (2005)
4. Thomson, M.G.A., Westland, S.: Color-imager calibration by parametric fitting of sensor responses. *Color Res. Appl.* 26, 442–449 (2001)
5. Imai, F.H., Quan, S., Rosen, M.R., Berns, R.S.: Digital camera filter design for colorimetric and spectral accuracy. In: Proceedings of the 3rd International Conference on Multispectral Color Science, Joensuu, Finland, pp. 23–26 (2001)
6. Hong, G., Ronnier Luo, M., Rhodes, P.A.: A Study of Digital Camera Colorimetric Characterisation Based on Polynomial Modelling. *Color Research & Application* 26(1), 76–84 (2001)
7. Cheung, V., Westland, S., Li, C.J., Hardeberg, J., Connah, D.: Characterization of trichromatic color cameras by usring a new multispectral imaging technique. *J. Opt. Soc. Am. A* 22(7) (2005)
8. Cheung, V., Westland, S., Thomson, M.G.A.: Accurate estimation of the non-linearity of input-output response for color cameras. *Color Res. Appl.* 29, 406–412 (2004)
9. Barnard, K., Funt, B.: Camera characterization for color research (2001),
http://www.CS.Berkeley.EDU/~kobus/research/publications/camera_characterization/index.html

Silverlight-Based Distance Teaching Application

Jin Wang

Guilin University of Electronic Technology
Teaching Practice & Experiment Department, China
access_wj@163.com

Abstract. This paper introduces the concept and characteristics of Silverlight technology, and its future development trend. Briefly presents the advantage of Slivelight technology in the application of distance learning interaction, and the realizing of a real teaching scene as well as real-time interactive experience in distance learning is discussed. In network teaching, Silverlight technology gives a great advantage for improving the interaction effects.

Keywords: Silverlight technology, interactive, experience.

1 Introduction

In the distance learning system, Web-based network teaching platform for its rich feature, structural integrity, quickly become the main carrier of the online education. The open learning conditions, rich learning resources enabling the network teaching has the incomparable advantages as the traditional teaching methods. However, in the online teaching, with the increasingly high demand for sense of interaction and experience, the traditional web network teaching which is data-centric, manifested in the form of HTML, the presentation layer can't facilitate the integration of different media forms, thus would cause students a higher, all-round experience requirement is difficult to meet, leading to a poor ability to interact, and the students' participation is not high.

This article focuses on the RIA solutions of the rapid development of multimedia vector animation technology – Silverlight. It makes full use of client computing power, with the aid of a client script engine, could put server's operation move to the client in part, and ensures the fast response and convenient communication. The application of Silverlight technology in the distance learning system could change the pattern of interaction about distance learners and teaching system, improve the interface quality of the teaching system, simultaneously offer a good experience for learners, and thereby enhancing the efficiency and quality of the learners.

2 Silverlight Technology

Silverlight is a cross-browser, cross-platform plug-in which Microsoft newly released, and it is the most important piece which develops the RIA technology, bringing the

media experience and rich interactive applications for the next generation based on the .net. Silverlight provides a flexible programming model, and could easily be integrated into the existing network applications programs. Silverlight provides high-quality video information for major browsers in fast, low-cost delivery running on Mac or Windows. As part of the strategy of the Microsoft .NET 3.0, Silverlight is to compete with the Flash of Adobe company, and will become the most expressive media-rich application architecture of the network.

Silverlight supports multiple operating systems of Microsoft and Apple. Such as Microsoft Windows XP SP2, Microsoft Windows Vista, Microsoft Windows 7, Microsoft Windows Server2003, and Apple MacOS X10.4.8 above and OS X Leopard operating system. For Linux system, Silverlight Lunix version is called the Moonlight.

Silverlight support IE, Firefox and other popular browsing device, including IE6, IE7, Firefox 1.5 above (Mac OS and Windows), Safari2.0 brower and Operabrower, etc.

3 Silverlight Application in the Distance Learning System

At present, the web-based LSM (Learning Management System) learning management system developed very rapidly. However, from the application perspective of the specific learning management system, it still has the problems that the expression of study contents, and the interaction force between the learners and teachers are all insufficient.

From the view of learning contents' expression, the presentation of contents based on Web learning management system are mainly in HTML form as the foundation, and combined with PPT and animation form to realize. These information performance technologies have their own deficiencies, and it is difficult to use a single information display technology to organize flexible learning support and to realize different learning scenarios. The web learning system realized by integrated applications based on these technologies, would inevitably appear issues, such as complex uses, organize difficult, lack of flexibility, etc.

From the view of interaction force between the learners and teaching, research shows that the network teaching lack of an intense atmosphere of emotional communication between the teachers and students like the traditional teaching. That's easily lead to network learning disabilities. These obstacles in addition to the learner's own personality and other factors, learning environment of the network platform also be related to. In the network learning process, interaction is the central link which realizes the teaching information transmitting, feedback, and finally form the knowledge, skills and emotion in the entire distance teaching process.

In order to solve above problems about network learning management system, it's necessary to establish a virtual classroom under the network environment. We should provide learning environment similar to the real classroom teaching, and make the network learners return to the classroom, feel the scene and atmosphere of the classroom teaching. Silverlight technology running on the Internet is not only able to

completely meet the requirement of this interaction, but also the perfect interactive experience.

3.1 Powerful Digital Ink Function

Silverlight built-in digital ink function. By calling the InkPresenter object inherited from the Canvas object, the normative area of the screen which could depict and show the digital ink, is used to memory and display the path which the user's mouse has passed, and provide additional support, including the digital pen, input device such as a touch screen, etc. At present, the Digital ink technology is widely applied to the tablet PC or PDA (Personal Digital Assistant), and it is a kind of technology to capture the handwriting on the computer. Reference handwritten recognition technology simultaneously, it also beyond lots of its limitations. It is silverlight this feature that enables us writing pages becomes as simple as holding a light pen to write on the touch screen, replacing the trouble and inconvenience of the keyboard input.

This function brings new experience for interaction in distance learning, and makes the interaction become simpler and convenient, improving the effectiveness and efficiency of the interaction.

3.2 Build an E-Learning Whiteboard of Full Range Interactive Features

Teaching whiteboard is an important teaching tool in distance learning. Its main function is to realize real-time communication and discussion that between the teachers and students, the students with students and the students with learning resources. In the teaching process, by the whiteboard that teachers could expound the content and put forward the question, and the students could also put forward problems through the white board, view the speech of teachers and other students, and achieve real-time online discussions and exchange.

The Silverlight-based e-learning whiteboard mainly achieved by XAML(eXtensible Application Markup Language) + WCF (Windows Communication Foundation) technology. In the presentation layer, using XAML language to do the drawing of whiteboard graphical interface, and provides flexible various interface control elements for interface. These control elements can be well combined with data model, so gives the user quick and convenient interactive experience. And the feedback of user's operation is through the WCF. The WCF programming model puts the Web service, .NET remote technology, distributed transactions and message queues into a single service-oriented programming model, realize the distributed computing that of the true sense, thus allows more people to use the whiteboard to draw pictures, send text messages at the same time, and these movements will be reflected in each user's screen. Therefore, the Whiteboard constructed by the Silverlight technology make the communication and discussion that between the teachers and students, students and students more efficient and convenient. And bring a full range of interactive experiences for students and teachers.

3.3 Perfect Video and Animation Broadcast

At present, due to the poor speed, when we play the educational video and multimedia courseware in distance learning, will often cause the video broadcast midway stopping or the player on and off, and the broadcast quality is not clear. This will greatly reduce the learners' learning motivation, and affects the positive interaction between the learners and multimedia learning resources, but the perfect effect of video and animation broadcast of silverlight is a good solution.

Silverlight support VC-1/H.2G4 / Audio Coding (AAC), audio, live or VOD Internet Information Services7.0 (IIS7Smooth Streaming), full HD (720p and above the resolution) playback, as well as support for the more scalable decoder, giving the users with rich, full screen, and no pauses in media experience. Meantime, through the Windows Media technology which has won the Emmy Awards, Silverlight could reduce the flow to 46%, and is compatible with the existing flow configuration of Windows Media. If adopt next-generation IIS media pack, which based on Windows Server system, the flow rate will decline further. Therefore, video player software based on Silverlight technology will not affect the playback quality when transfer or play lots of content. With the development of multimedia technology, distance learning's demand for high-definition network video is more and more high. Silverlight can also play high-definition version of format H.264. When read the data and renew the appearance, Silverlight could adopt no refresh page way to update and do not affect the user's operation.

Silverlight support for animation, video, audio, two-way data communications, and could create a complex form, so when the students get a good learning experience, can also maximize participation in interactive.

3.4 3D Modeling Constructs the Virtual Interactive Scene

The latest version silverlight3 will provide the full support of 3D modeling, so in the Silverlight application, any future developers could introduce 3DMAX, Maya, etc, modeling tools to create 3D models. Under the normal circumstances, a simple model directly uses XAML markup language to design, but a complex model uses MAYA, PRO/E application software to shape. First, entered into 3DMAX, through conversion software provided by the Microsoft would be output in XAML file, and then generates a virtual scene in the browser through Silverlight. Therefore, due to the emergence of this technology, it becomes possible that we could develop the education game with real-time interactive 3D, build virtual laboratory, and simulate the real teaching scenarios in the remote teaching based on WEB. Through 3D virtual teaching scene, we build virtual discussion scene, virtual team learning scenarios, virtual-learning resources, and providing a realistic teaching environment for learners. All of these, cannot only make the learners enjoy the freedom learning and practice space, but also do various of real-time communication and cooperation, and bring the perfect learning and interactive experience to the learner.

4 Summary

To sum up, we have known the flexibility and interactive applications of Silverlight, as well as its development out of the infinite network charm. Silverlight as RIA application development technology, using its characteristics such as rich media elements presenting ability, the separation of animation and data content, and the ability that supports various ways of network communication with the clients, meantime, the interface control logic is transferred to the client to achieve, that would be able to develop rich expressive and high interactive demanding applications. Through comprehensive utilization Silverlight's various technical characteristics, by single technology, network distance learning could realize a web real-time classroom that has strong user experience and high realism. It solves the problem that the user communication scene experience is insufficient in a realistic learning system, and will have a wide application value in network teaching. I believe that with the scientific progress and the increasingly intense competition, Silverlight will be more powerful and perfect, and eventually become the leader in the relevant technical fields.

References

1. Cheng, G.-X., Hu, S.-Q.: Research and Implementation of an Interactive Learing RIA Platform Based on Silverlight. Computer Engineering & Science 32, 23–26 (2010)
2. Peng, S.U.: Magic Silverligh, pp. 40–62. Publishing House of Electronics Industry, Beijing (2008)
3. Li, H.-J.: Silverlight 2 Perfect journey, pp. 475–486. Publishing House of Electronics Industry, Beijing (2009)
4. Cheng, G.-X., Hu, S.-Q.: Design and implementation of network Virtual Teaching System Based on Silverlight. China Educational Technology 293, 126–129 (2011)
5. Zhang, Y.: The realizing of University students' comprehensive management system based on Silverlight and WCF technology. Software Engineering. Yunnan University, Yunnan (2010)
6. Wang, N.: Design and Implementation of Ontology Editor Based on Silverlight Technology. Computer Software and Theory. Ocean University of China, Shandong (2010)
7. Sun, X.-Q.: Silverlight-based application architecture design. Science & Technology Information 34, 248 (2010)
8. Xiong, Y.-L.: Research and Application of RIA Based on Silverlight. Computer Software and Theory. Chongqing University, Chongqing (2008)
9. Sun, W.: The Application of RIA to the E-Learning Platform. Modern Distance Education Research 03, 84–88 (2011)
10. Zhong, L.: Evolving from Teaching Platform To Learning Platform: Software Engineering E-Learning Platform as an Example. Modern Educational Technology 21, 103–107 (2011)

Uncorrelated Discriminant Isometric Projection for Face Recognition

Bo Ge^{1,2}, Yanling Shao², and Yunxing Shu¹

¹ Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, 471023 Luoyang, Henan P.R. China

² Henan University of Science and Technology, 471003 Luoyang, Henan P.R. China
`{wgq2211,sweetsylgy,65929101}@163.com`

Abstract. Feature extraction is a crucial step for face recognition. In this paper, based on Isometric Projections (IsoP), a new feature extraction method called Uncorrelated Discriminant Isometric Projections (UDIsoP) is proposed for face recognition. The aim of UDIsop is to preserve the within-class geometry structure by taking into account the class label information. Moreover, the features extracted via UDIsop are statistically uncorrelated with minimum redundancy, which is desirable for many pattern analysis applications. Experiment results on the publicly available ORL and Yale face databases show that the proposed UDIsop approach provides a better representation of the data and achieves much higher recognition accuracy.

Keywords: Feature extraction, Manifold learning, Isometric Projections, Uncorrelated constraint, Face Recognition.

1 Introduction

Face recognition [1], the identification of individuals in video supervision system, target tracking and recognition system, has gained significant attention, in which feature extraction is a crucial step. The two most well-known feature extraction methods are Principle Component Analysis (PCA) [2], [3] and Linear Discriminant Analysis(LDA) [4]. Affected by many factors such as expression illumination and pose, the face images likely reside on a nonlinear face manifold [5], However, both PCA and LDA only effectively discover the global linear structure.

Recently, manifold learning, based on local geometry analysis, has drawn much attention. Unlike PCA and LDA, which aim to preserve the global Euclidean structure of the data space, manifold learning algorithms aim to preserve the inherent manifold structure. The most representative algorithms include Locally Linear Embedding (LLE) [6], ISOMAP [7], Laplacian Eigenmap [8], Local Tangent Space Alignment (LTSA) [9], Local Coordinates Alignment (LCA) [10], Local Spline Embedding (LSE) [11], and so on. However, these methods are defined only on the training set, and the issue of how to map new test data remains difficult. Due to this problem, linear manifold embedding methods are proposed, such as Isometric Projections

(IsoP) [12], Neighborhood Preserving Projections(NPP) [13] and Locality Preserving Projections (LPP) [14] which can find the mapping on the whole data space. Note that, although IsoP is successful in mapping training samples and testing samples into a low-dimensional space, it nevertheless suffers from limitations. Firstly, It deemphasizes discriminant information that is important for recognition task. Secondly, the basis vectors obtained by IsoP are statistically correlated, and so the extracted features contain redundancy, which may distort the distribution of the features and even dramatically degrade recognition performance [15].

In this paper, a novel feature extraction method called Uncorrelated Discriminant Isometric Projections (UDIsoP) is proposed. Based on IsoP, UDIsoP utilizes the class label information to preserve the within-class geometric structure and achieves good discriminant ability. On the other hand, the method obtains statistically uncorrelated features with minimum redundancy by introducing a simple uncorrelated constraint on the computation of the basis vectors[16],[17].

2 A Brief Review of Isometric Projections

Isometric Projections (IsoP) is a linear approximation of ISOMAP. Let $X = [x_1, x_2, \dots, x_N]$ be a set of original data in which R^D is the high-dimensional space, M^p is the embedded manifold in the low-dimensional space ($p \leq D$). The aim of this algorithm is to seek a transformation matrix V , and maps these data points to new points $Y = [y_1, y_2, \dots, y_N] = V^T X$ in $d(d \ll D)$ -dimensional space.

In order to preserve the intrinsic geometrical structure of the data manifold, seek a function f which satisfies the following requirement by considering the projection $f : R^D \rightarrow R^d (d \ll D)$, such that:

$$d_{M^p}(x, y) = d_{R^d}(f(x), f(y)) \quad (1)$$

Therefore, we can design the following function:

$$f^{opt} = \min \sum_{i,j} (d_{M^p}(x_i, x_j) - d_{R^d}(y_i, y_j))^2 \quad (2)$$

In practical applications, the distance on the manifold $d_{M^p}(x_i, x_j)$ can not be obtained directly for the underlying manifold M^p is unknown. Under this situation, the nearest-path method is utilized to get $d_{M^p}(x_i, x_j)$ which is the computing approximation of distance on the neighborhood graph actually.

Let D_{ij} be the distance matrix between x_i and x_j . Define $S_{ij} = D_{ij}^2$ and $H = I - (\frac{1}{N})ee^T$ where I is the identity matrix and e is the vector of all ones. It can be shown that $\tau(D) = -(\frac{1}{2})HSH$ is the inner product matrix.

The objective function of Eq.(2) is equivalent to:

$$\|\tau(D_G) - \tau(D_Y)\|_{L^2} \quad (3)$$

where $\|A\|_{L^2} = \sqrt{\sum_{i,j} A_{i,j}^2}$. Considering the function of $f(x) = V^T x$, Eq.(2) can be converted to:

$$\begin{aligned} V^* &= \min \|\tau(D_G) - \tau(D_Y)\|^2 \\ &= \min \|\tau(D_G) - X^T V V^T X\|^2 \end{aligned} \quad (4)$$

After derivation, Eq.(4) can be rewritten as follows:

$$\begin{cases} \arg \max_V V^T X \tau(D_G) X^T V \\ s.t. V X X^T V = 1 \end{cases} \quad (5)$$

According to Lagrange multiplier method, Eq.(5) becomes:

$$V^T X \tau(D_G) X^T V = \lambda X X^T V \quad (6)$$

3 Uncorrelated Discriminant Isometric Projection

IsoP is an unsupervised feature extraction method. As mentioned above, discriminant information is important for recognition problem. For this reason, we modify IsoP with class label information and yield Uncorrelated Discriminant Isometric Projection (UDIsoP). The goal of UDIsoP is to preserve the within-class neighborhood geometry structure by taking into account class information.

In order to embody the discriminant information on the manifold, we constructed a discriminant weight w_{ij} , which is used to compact points x_i and x_j if they share the same class label, and expand x_i and x_j if they belong to different classes. Then we have:

$$\min \sum_{i,j} (w_{ij} d_M(x_i, x_j) - d_E(y_i, y_j))^2 \quad (7)$$

Note that, an informative factor represented in discriminant matrix W , which is determined by class label information $l(i, j)$ and the geodesic distance $d_M(x_i, x_j)$ on M between points i and j . We also considered the compacting and expanding factors. So we have:

$$w_{ij} = f(d_M(x_i, x_j), l(i, j), P_c, P_e) \quad (8)$$

$l(i, j)$ is a label function and gives the correlation of points i and j , defined as follows:

$$l(i, j) = \begin{cases} 1, & \text{if } x_i, x_j \text{ share the same label} \\ 0, & \text{if } x_i, x_j \text{ have different labels} \end{cases} \quad (9)$$

where P_c, P_e are the compacting and expanding parameters, respectively. They are reflecting the degrees of compacting and expanding.

Then, Eq.(7) can be converted to a matrix form:

$$\min \|W \bullet H(D_G) - H(D_Y)\|_2 \quad (10)$$

Let $A = W \bullet H(D_G)$, which $[A]_{ij} = [W]_{ij} \times [H(D_G)]_{ij}$

So far, the objective function can be rewritten as :

$$\arg \max V^T X A X^T V \quad (11)$$

Next, the statistically uncorrelated constrain is considered. Assumed any two different components y_i and y_j ($j \neq i$) of the extracted feature $y = V^T x$ are uncorrelated; then,

$$E[(y_i - E(y_i))(y_j - E(y_j))] = v_i^T S_t v_j = 0 \quad (12)$$

where v_i and v_j are two different columns of the matrix V . S_t is the total scatter matrix of the training set. Besides, v_i should be normalized. Let v_i satisfy

$$v_i^T S_t v_i = 1. \quad (13)$$

Note that the above normalization is only for simplifying the computation. Once we get the optimal solutions, we can renormalize them to get orthonormal basis vectors. Then, from Eq.(12) and Eq.(13), we can get

$$V^T S_t V = I \quad (14)$$

As a result, UDIsoP can be formulated as the following constrained maximization problem:

$$\max_{V^T S_t V = I} \text{tr}(V^T X A X^T V) \quad (15)$$

Using Lagrange multiplier, let:

$$L(V, \lambda) = V^T X A X^T V - \lambda(V^T S_t V - I)$$

Then

$$\frac{\partial L(V, \lambda)}{\partial V} = X A X^T V - \lambda S_t V = 0$$

Thus, the constrained maximization problem can be converted to the following generalized eigenvalue problem:

$$XAX^T V = \lambda S_t V. \quad (16)$$

Let the column vector v_1, v_2, \dots, v_d be the solutions of Eq.(16) order according to their eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$. Thus, the embedding is given as follows:

$$\begin{aligned} X &\rightarrow Y = V^T X \\ V &= [v_1, v_2, \dots, v_d] \end{aligned} \quad (17)$$

4 Experiment Results

To demonstrate the effectiveness of the proposed UDIsop approach, experiments were done on the ORL database and Yale database. The ORL database contains images from 40 individuals, each providing ten different images. The images are captured at different times and have different variations, including expressions and facial details. The Yale database contains 165 face images of 15 individuals. There are 11 face images per individual, and these images demonstrate variations in facial expression (happy, normal, sad, sleepy, surprised, and winking) and lighting condition (center-light, left-light, right-light).

For the purpose of computational efficiency, all images in ORL database and Yale database are resized to 32×32 pixels with 256 gray levels per pixel. Fig.1 and Fig.2 show the cropped and resized image samples of one individual in ORL database and Yale database respectively.

To evaluate the efficiency of the UDIsop algorithm, we compare the accurate recognition rate of the proposed method with that of other methods such as Eigenfaces (PCA), Fisherfaces (LDA), NPP, LPP and IsoP.



Fig. 1. Sample face images from the ORL database



Fig. 2. Sample face images from the Yale database

4.1 Experiments on ORL Database

In the experiments, n (n varies from 4 to 6) face image samples are randomly selected from the image gallery of each individual to form the training sample. The remaining are used for testing. For each given n , we performed 10 times by randomly choosing different training and testing sets. Finally, a nearest-neighbor classifier is employed for classification.

Table 1 contains comparative analysis of the obtained maximal average recognition rate, the corresponding standard deviations and the reduced dimensions for six algorithms (PCA, LDA, NPP, LPP, IsoP and UDIsop) on the ORL database. From Table 1, we can see that the performance of the proposed UDIsop algorithm outperform other algorithms. It demonstrates that the performance is improved because UDIsop takes into account the class label information to preserve the within-class geometric structure and the extracted features are uncorrelated. Due to the two aspects, UDIsop achieves better performance.

Table 1. Recognition accuracy (%) comparison on ORL face database

Method	4Train	5Train	6Train
PCA	85.76 \pm 2.16(159)	88.35 \pm 1.85(199)	90.63 \pm 2.01(239)
LDA	91.29 \pm 2.55(39)	93.56 \pm 1.48(39)	94.05 \pm 1.90(39)
NPP	91.35 \pm 2.21(39)	93.55 \pm 1.28(39)	94.69 \pm 2.16(42)
LPP	91.53 \pm 2.15(40)	94.20 \pm 1.53(39)	94.62 \pm 1.91(39)
IsoP	91.99 \pm 2.24(39)	93.56 \pm 1.72(39)	94.85 \pm 2.12(39)
UDIsop	92.12 \pm 1.81(39)	95.12 \pm 2.02(39)	96.06 \pm 2.18(39)

4.2 Experiments on Yale Database

Similar to the experimental procedure on ORL face database, for each individual, n ($= 4, 5, 6$) face image samples are randomly selected for training and the rest are used for testing. For each given n , we average the results over 10 random splits. The experiment results include the maximal average recognition rate, the corresponding standard deviations (std) and the reduced dimensions which are shown in Table 2.

Table 2. Recognition accuracy (%) comparison on Yale face database

Method	4Train	5Train	6Train
PCA	53.44 \pm 2.23(59)	58.69 \pm 3.61(74)	59.72 \pm 2.63(89)
LDA	67.68 \pm 3.81(14)	74.12 \pm 4.01(14)	75.06 \pm 4.57(14)
NPP	72.04 \pm 3.45(14)	76.53 \pm 4.36(14)	78.13 \pm 3.34(22)
LPP	71.90 \pm 2.88(14)	75.36 \pm 3.61(14)	78.80 \pm 4.13(15)
IsoP	72.68 \pm 3.23(14)	76.57 \pm 4.22(14)	79.33 \pm 3.84(15)
UDIsop	73.94 \pm 3.14(14)	78.26 \pm 4.18(14)	81.06 \pm 3.95(14)

From Table 2, it is obvious that UDIsoP performs the best for each n . The table indicates that UDIsoP has more discrimination power than other methods and that the statistically uncorrelated property of extracted feature is helpful for improving recognition performance.

5 Conclusions and Future Work

An Uncorrelated Discriminant Isometric Projection algorithm (UDIsoP) is proposed in this paper. UDIsoP takes class information of training data into account when constructing the weighted neighborhood graph, thus it can preserve the within-class neighborhood geometry structure. Moreover, an uncorrelated constraint is introduced to generate statistically uncorrelated features. Experimental results on ORL and Yale face databases indicate UDIsoP has more discriminative power than PCA, LDA, NPP, LPP and IsoP. But UDIsoP algorithm is still linear. In future work, we will extend UDIsoP to nonlinear form by kernel trick.

Acknowledgments. This work is Supported by Foundation of He'nan Educational Committee (number 2010A520028).

References

1. Zhao, W., Chellappa, R., Phillips, P.J., et al.: Face recognition: a literature survey. ACM Computer Surveys 35(4), 399–458 (2003)
2. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (1986)
3. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience 3(1), 71–86 (1991)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)
5. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face Recognition Using Laplacianfaces. IEEE. Trans. Pattern Analysis and Machine Intelligence 27(3), 328–340 (2005)
6. Roweis, S., Saul, L.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science 290, 2323–2326 (2000)
7. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290, 2319–2323 (2000)
8. Belkin, M., Niyogi, P.: Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In: Advances in Neural Information Processing Systems 14, Vancouver, British Columbia, Canada (2002)
9. Zhang, Z., Zha, H.: Principle Manifolds and Nonlinear Dimensionality Reduction via Tangent Space Alignment. SIAM J. Sci. Compute 26(1), 313–338 (2004)
10. Zhang, T.H., Li, X.L., Tao, D.C., Yang, J.: Local Coordinates Alignment (LCA): a Novel Method for Manifold Learning. International Journal of Pattern Recognition and Artificial Intelligence 22(4), 667–690 (2008)
11. Xiang, S.M., Nie, F.P., Zhang, C.S., Zhang, C.X.: Nonlinear Dimensionality Reduction with Local Spline Embedding. IEEE Trans. on Knowledge and Data Engineering 21(9), 1285–1298 (2009)

12. Cai, D., He, X.F., Han, J.W.: Isometric projection. In: Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI 2007), Vancouver, Canada, July 22–26, pp. 528–533. AAAIPress, Menlo Park (2007)
13. Pang, Y., Zhang, L., Liu, Z., Yu, N., Li, H.: Neighborhood Preserving Projections (NPP): A Novel Linear Dimension Reduction Method. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 117–125. Springer, Heidelberg (2005)
14. He, X., Niyogi, P.: Locality Preserving Projections. In: Proc. Advances in Neural Information Processing System Conf. (2003)
15. Sun, S.Y., Zhao, H.T., Yang, H.J.: Discriminant Uncorrelated Locality Preserving Projection. In: Proc. of International Conference on Image and Signal Processing, pp. 1849–4852 (2010)
16. Wang, G.Q., Zhang, W.J., Liu, D.T.: Discriminant Uncorrelated Neighborhood Preserving Projections. Journal of Computational Information Systems 6(4), 1027–1035 (2010)
17. Liu, H., Yang, J.A., Wang, Y.: Feature Extraction of Acoustic Targets Using Uncorrelated Neighborhood Preserving Discriminant Projections. Journal of Computational Information Systems 6(8), 2691–2698 (2010)

Design and Realization of Multimedia Animation Based on Flash

Shouping Wang¹ and Zhenbing Wang²

¹ International College, Qingdao University, Qingdao City, Shandong 266071, China

² Network and Modern Educational Technology Center, Binzhou College,
Shandong 256600, China

wsp1900@hotmail.com, bwzwb@163.com

Abstract. Flash-based multimedia animation has been proven to be widely applied in education and teaching fields in the appropriate form. Taking a series of multimedia animations of Shandong Science and Technology Museum as an example, the paper describes the meeting of multimedia animation design and education thought of Mayer. On the basis of multimedia learning theory, multimedia animation design should pay attention to integrate resources, display information with multimedia means, avoid cognitive load, enhance interaction and promote learners' active process. In addition, there are essential differences between ActionScript3 and former ActionScript2. The paper also discusses how to write code to realize the common function in multimedia animation in ActionScript3.

Keywords: Multimedia animation, Multimedia learning, Flash, AS3.0.

1 Introduce

With the development of computer technology, multimedia animation that presents information in various forms such as text, static graphics, sound, animation, video, and so on, is more and more widely used in education and teaching. At present, however, lack of the guidance of scientific theories and research of learner, paying too much attention to technical performance, multimedia animation looks beautiful, but has poor effect in education and teaching. Taking a series of multimedia animation coming from provincial subject that the author has taken part in as an example, the paper introduces how to combine Mayer's multimedia learning theory with design practice, and how to write codes in ActionScript3 to realize the common function in multimedia animation.

2 Mayer's Theory of Multimedia Learning

Richard E. Mayer is the famous contemporary education psychologist, experimental psychologist, the pioneer and leader of international cognitive psychology research in

multimedia learning[1]. His outstanding contributions are to propose multimedia cognitive theory and seven principles in multimedia design.

Drawing on dual coding theory, cognitive load theory and constructivism learning theory, Mayer presents multimedia learning theory [2,3]. The theory holds that multimedia teaching information is shown in words and pictures. Through the corresponding verbal and visual channel, words and pictures are paid attention to, then processed in different channel and integrated between the visual and verbal mental models, with learners' prior knowledge in long-term memory, in order to complete meaningful learning.

Mayer thinks that education psychology research should take into account both theory and practice [4]. Since 1990, Mayer and his colleagues made more than 100 experiments. By the analysis of experimental results, based on cognitive theory of multimedia learning, Mayer puts forward seven multimedia design principles: multimedia principle, spatial contiguity principle, temporal contiguity principle, coherence principle, modality principle, redundancy principle and individual differences principle.

3 Design Strategy of Multimedia Animation

The most important characteristic in Science and Technology Museum is exhibition education. With the development of information technology, the display form has been developed to multimedia technology from traditional sound, light and electricity. Science and Technology Museum and Multimedia Research Center of Shandong Normal University, present a series of multimedia animations in 2003 that early realized multimedia display. Facing children and teenagers, animations are divided into seven series such as the big bang, ocean--the origin of life, biological evolution and mass extinction, and so on. Developed with Flash, scientific knowledge becomes vivid and attractive. The exhibits are thought highly by teenagers and experts.

The author has participated in the design and felt deeply it's important to combine Mayer's multimedia cognitive theory and design principles with multimedia animation design. Now the paper will discuss it in such respects as content choice, information presentation methods, and so on.

3.1 Resource Integration Strategy

Mayer's multimedia cognitive theory holds that the best learning is to construct and integrate model, finally forming long-term memory in the brain. In effective multimedia teaching, the cognitive mechanism of learners involves such basic cognitive activities as perception, attention, selection, organization and integration of information, and its core and ultimate aim is to form long-term memory [5]. The learning results are divided into no learning, rote learning, and meaningful learning by Mayer in which the latter is efficient learning which constructs knowledge model and forms long-term memory[6].

Integration of a series of multimedia animations in Shandong Science and Technology Museum is embodied in around the specific theme, information is presented hierarchically. For example, in animation "biological evolution and mass extinction", around the theme of biological evolution and extinction, the directory structure is shown in Figure 1. By clicking on a button, viewers can enter into the home page and different levels of sub page. With systematic content design and distinct layers, every animation is conducive for viewers to integrated new knowledge continuously, with his previous experience, and form long-term memory.

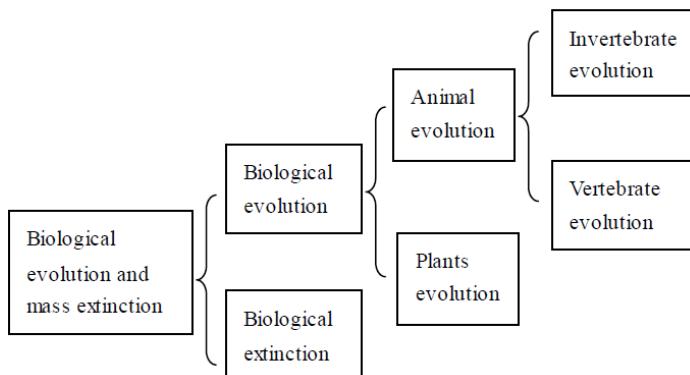


Fig. 1. Directory structure of home page in multimedia animation "biological evolution and mass extinction"

3.2 Multimedia Strategy

Unlike previous children, now primary and middle school students live in the digital world and technology has become an indispensable part in daily life [7]. With the rapid development of information technology and the understanding of brain's information processing, it is probable to promote the students' learning by multimedia significantly.

Mayer's multimedia cognitive principles hold that it is better to present information in words and pictures than in only words. According to dual coding hypothesis, human processes information in different visual and verbal channels that are relatively independent. Learners are more likely to be able to make connections between corresponding visual and verbal representations in working memory when the animation and narration are coordinated in time.

According to multimedia cognitive principles, a series of multimedia animations of Shandong Science and Technology Museum present information in pictures and narration. Based on information content, it involves various kinds of picture forms, such as static pictures, Flash animation, video or any combination of them.

3.3 Interaction Strategy

In Mayer's opinion, multimedia learning will be more effective when learning material is interactive and in control of learners [8]. Not all students study at the same speed. If learners can control the speed of presentation, it has better results.

A series of multimedia animations of Shandong Science and Technology Museum take sector-by-sector principle. As far as the biological extinction is concerned, it has three branches: five biological extinction, extinction being carried on, protecting dependent friends of mankind. With five biological extinctions, it is divided into five parts: the first, second, third, fourth, and fifth extinction. Learners can click to play the animation and narration of each part. To avoid the overload, each sector contains about two or three sentences. The interaction strategy is also manifested in animation playback during which viewers can control the playback speed by clicking start, pause, forward and backward.

3.4 Avoid Cognitive Load

According to cognitive load theory, learners' cognitive load is divided into internal cognitive load, external cognitive load, germane cognitive load. External cognitive load caused by poor design refers to the effort to process teaching material badly designed that has no contribution to study [9]. To reduce external cognitive load, information presentation of a series of multimedia animations keep consistent in design styles of each directory page and button.

The redundancy principle which Mayer puts forward is that it is better to present animation and narration than animation, narration, and on-screen text [10]. With a series of multimedia animation, presentation takes the model of animation and narration, only key words or key phrases occasionally appearing on the screen. However, some designers think learners have different learning preferences. If the same information is presented via each mode, we provide more modes allowing for more students to choose their preferred mode. However, adding on-screen text can create a split-attention effect in which students must look both at the animation and text, thereby missing out some presented material. When visual working memory is overloaded, there is less cognitive energy to build connections between visual and verbal representations [11].

4 Writing Action Script of Multimedia Animation

Action script in a series of multimedia animations of Shandong Science and Technology Museum mainly controls jumping between different pages and playing narration. Taking biological mass extinction in "evolution and biological extinction" as example, the paper introduces how to write common action script. It's worth noting that the contents of home page, such as five times biological extinctions, biological extinction still on, protection of human interdependence friends are placed in the first frame, the second frame, the third frame, the fourth frame in each layer of timeline.

4.1 Jumping between Different Pages

Buttons for five biological extinctions, extinction still on, protection of human dependent friends, and return are named respectively miejue_btn, jixu_btn, baohu_btn, fanhui_btn in property panel. Add a new layer in timeline, and enter the following code in action script panel:

```
stop();
function plymj (event:MouseEvent):void {
    gotoAndPlay(2);
}
function plyjx (event:MouseEvent):void {
    gotoAndPlay(3);
}
function plybh (event:MouseEvent):void {
    gotoAndPlay(4);
}
miejue_btn.addEventListener(MouseEvent.CLICK, plymj);
jixu_btn.addEventListener(MouseEvent.CLICK, plyjx);
baohu_btn.addEventListener(MouseEvent.CLICK, plybh); [12]
```

Insert a blank key frame in the second frame of five biological mass extinctions page and enter the following code:

```
stop();
function plyml(event:MouseEvent):void{
    gotoAndPlay(1);
}
miejue_btn.addEventListener(MouseEvent.CLICK,plymj);
jixu_btn.addEventListener(MouseEvent.CLICK,plyjx);
baohu_btn.addEventListener(MouseEvent.CLICK,plybh);
fanhui_btn.addEventListener(MouseEvent.CLICK,plyml);
```

In the same way, enter code in other pages. It must be noted that the mouse listener should be added to different button.

4.2 Calling Corresponding Narration for Different Pages

To facilitate calling of the file, narration and Flash files should be placed in the same folder. Enter codes as follow in action panel of code layer of home page:

```
SoundMixer.stopAll();
var sml:Sound=new Sound();
var reqml:URLRequest =new URLRequest ("mulu.MP3");
sml.load(reqml);
sml.play(1); [13]
```

Enter codes for remaining pages in the same way, only defining instance and variable for different narration.

5 Conclusion

With various presentation forms, multimedia animation makes learners impressed, so it has been more and more widely used in education, training and other fields. However, lacking of the guidance of scientific theory, designers often set out from experience. Moreover, designers are entangled by the emergence of new ActionScript3. The paper discusses the interface design of multimedia animation and how to write code for common function in ActionScript3, hoping to provide useful reference for designers in practice.

Acknowledgements. Part of this work is supported by instruction research project from Qingdao University: the research and practice of animation instruction reform, contract JY1057.

References

1. Mayer, R.E., <http://www.psych.ucsb.edu/people/faculty/mayer/index.php>
2. Mayer, R.E., Moreno, R.: Aids to computer-based multimedia learning. *Learning and Instruction* 12, 107–119 (2002)
3. Mayer, R.E.: Multimedia aids to problem-solving transfer. *International Journal of Educational Research* 31, 611–623 (1999)
4. Mayer, R.E.: Old Advice for New Researchers. *Educational Psychology Review* 20(1), 19–28 (2008)
5. Mayer, R.E.: Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review* 8(4), 357–371 (1996)
6. SEG Research. Understanding Multimedia Learning: Integrating multimedia in the K-12 classroom. SEG Research Committee, USA (2011)
7. Mayer, R.E.: Cognitive theory and the design of multimedia instruction: An Example of the Two-way Street between Cognition and Instruction. *New Directions for Teaching and Learning* 89, 55–71 (2002)
8. Bannert, M.: Managing cognitive load-recent trends in cognitive load theory. *Learning and Instruction* 12, 139–1467 (2010)
9. Kirschner, P.A.: Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction* 12, 1–10 (2010)
10. Mousavi, S., Low, R., Sweller, J.: Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology* 87, 319–334 (2000)
11. Network of instruction, <http://bbs.jcwn.com/thread-259025-1-1.html>
12. Adobe, http://help.adobe.com/zh_CN/flash/cs/using/WSd60f23110762d6b883b18f10cb1fe1af6-7beca.html

The Ship Propeller Design System for Coupling of CAD, CAE and CAO

Jinfeng Huang

China Ship Development and Design Center, Wuhan, China 430063
howardhjf@163.com

Abstract. This paper presents an approach for tight coupling of parametric geometry design and associated CFD and CAE analysis. It is pointed out how simulation software can be technically integrated for the purpose of an automated process chain and efficient development of flow exposed products. In particular, the established process chain focuses on systematic studies and formal optimizations for which seamless interaction between CAD, CFD and CAE is required. The software SYSWARE-Framework is utilized for demonstration of the concept and an exemplary industrial application is described in detail. Ship propeller design system is an application platform which is based on flow, simulation, optimization and knowledge driven, mainly including propeller theory design, propeller automatic modeling, propeller hydrodynamics analysis and propeller structural strength analysis. This system can validate the possibility and efficiency of the technical method of propeller integrated design and analysis, lead and promote the creation of ship integrated design platform, and eventually integrate with the overall management platform in enterprise. The research of ship production will be efficient and effective by applying the platform.

Keywords: Propeller, Three-dimensional Model, Template Technique, CFD, CAE, RANS, Lifting Surface Theory.

1 Introduction

With information technology rapidly developing, ship propeller design mode is applied by CAD/CAE/CFD/CAO, etc. The ship design enterprises and the shipyards nowadays use the 3D design technology to create the 3D models of the ship propeller. Due to the complexity of the ship propeller, the simulation, analysis, optimization and evaluation of the design, which helps to determine whether the design is advantageous or not and predict whether the performance of the product could meet the demands or not. Therefore, it is necessary to simulate the functions and evaluate the performances of the product on the basis of the geometrical model in the designing process. Years before, due to the inefficiency in converting the data and integrating the information between the ship designing CAD system and the simulation analysis CAE software, it was unable to establish the connected relationship between the geometrical models and simulation analytical models so that the models could not be directly applied in the engineering analysis and calculation, this is called Information Island. Moreover,

the designers have to reestablish the simulation analytical model of the ship propeller in the analysis phase, which brings amount of repetitive work. Meanwhile, it is also not easy to ensure the consistency of the product model information, avoid the designing errors and implement the multi-discipline collaborative design and simulation. This is not beneficial for the comprehensive design optimization of the product.

Based on this background, the connected relationship between the geometrical model and simulation analytical model can be built with the integrated frame SYSWARE as the platform and the 3D geometrical model as the main model. The ship propeller integrated design system can be established with the template technology as the developing notion and with the success of the key technology of the data flow, control flow and process integration between the ship 3D design model and the simulation analytical model. Thus it is feasible to simulate the ship functions and performance, improve the designing efficiency and quality and achieve the goal of integrated designing optimization.

2 Basic Method

2.1 Related Model on Knowledge Based Engineering

Conception and technology of knowledge based engineering (KBE called) comes of 70 ages last century. In 1998, American put forward an idea which developed new generation CAX (CAD/CAE/CAM/CAO) based on KBE, Which lead CAX system into new developing phase. Artificial intelligence and CAX system were related by KBE, including knowledge acquiring, accumulating, combining, sharing, applying and improving, etc. Therefore, KBE is computer software system based on production model design, analysis and manufacture, which saves and deals with production model knowledge.

The ship propeller design initially establishes the 3D design model, also called geometrical model. Through the converting and processing of the geometrical model, it is effective to generate the ship propeller simulation model, perform analysis and calculation in the simulation software and analyze or evaluate the design result. So, the 3D design model is set as the main model while the simulation models of other disciplines which are derived from the main model are set as the related models. Once the main model is revised, the related models are revised simultaneously. The unified related models are the series and parallel connected networks of a serial of modules. The different models in the design proposal are coupled and related organically by building the data flow and control flow among the modules and establishing the unified related models that describe the entire design proposal. Therefore, when any local variation occurs in the design proposal, the rest is ensured to adjust to the variation automatically. This actually represents the process, method and data flow of the ship propeller design, so the key point in implementing the related models is to divide the unit design process and clearly define the ship design data flow. In each

module of the design process, the outward design parameters constitute parametric space of the transformed unified related models.

2.2 Template Technology

In order to implement the unified relation of the ship propeller models, the main measure is to establish an integrated design and analysis platform based on the digital ship on which the template technology is applied to fix the tools, software, design parameters and experiences and drives of processes and connections parameters are implemented through data flow and control flow.

In fact, any design or analysis process contains some regulations and methods which are the knowledge and experiences of using software tools. Though the objects the software operates are different from each other, the regulations and methods remain the same. The notion of this paper is to generalize and conclude the regulations, pack them as the template and perform the design and analysis of the product with the template.

According to the inner packed software and regulations, the template could accomplish the design, modeling, analysis, simulation and other different works and describe the various fundamental elements. It is feasible to use the template to establish the product design proposal as putting up the blocks. In order to ensure the connectivity of the design proposal, the system establishes the unified related model describing the entire design proposal via establishing the data flow and control flow among the modules and then connects different templates in the proposal organically. In such method, when any local variation occurs, the rest parts of the design proposal can adjust to the variation automatically. Therefore, the unified related model actually is the template integration linked by the data flow and control flow, which contains the entire design models, analytical models, analytical results and design analytical reports.

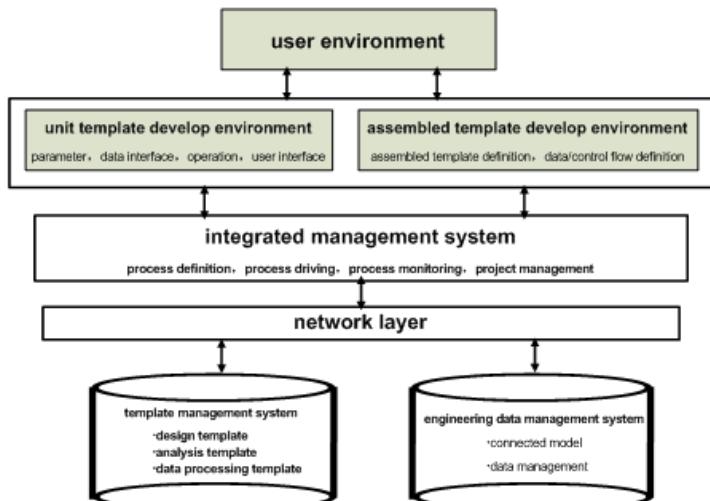


Fig. 1. The system hierarchy

3 The Structure and Component of the System

3.1 The Structure

The integrated ship general design platform is an open platform which is mainly used in integrating the software such as CAD, CAE, CAO, CAX etc and packing the applicative knowledge and experience of the software. It is in a unified environment that CAD design, modeling, analysis, optimization, data management and collaborative management are performed. Due to its functional characteristics, it is very suitable for the ship general design which is based on 3D design, adopts large amount of CAD and CAE technology, covers many disciplines in the design process and has very complicated design process.

The system hierarchy is shown in Fig.1 as below.

3.2 Compositions and Functions

User Environment

Because the product design, modeling and analysis involve abundant 3D models, the system, as propeller design system that directly opens to the designer, needs to visualize the models. For the sake of this, the CAD can be directly embedded in the system as graphics display then the geometry, grids and analytical results are simultaneously displayed in CAD. Moreover, in the CAD environment, the users are able to manipulate the design object generated from the template. User environment also takes responsibility of user management and user permission setting. This is not specified in the paper.

Unit Template Development Environment

The unit template development environment is the packing system which is integrated in CAD-environment-based design and analysis model or application software. Unit template development environment integrates and packs all kinds of design and analysis model or application software, defines and customizes the application interface which is directly for specific tasks and drives the design, analysis, modeling in simulation software, solution and post-process in the background. Thus it significantly lowers the degree of difficulty in using the professional software and decreases the repetitive workload in design and analysis. Meanwhile, the fixation and management of the software knowledge and experience are effectively implemented.

(1)Parameter table: this is the source of the template data, also the data core of all operations. Parameter table contains miscellaneous parameter types (integer, real, string, file, integer array, real array, string array). The users ought to input the parameters from the parameter table interface in which the real type parameter and file type parameter have their own interface separately.

(2)Input and output table: this is the template packing interface data and the main carrier of data flow transfer when executing the template. The input table, as the data resource of complex object, is responsible for receiving the data reference from the upstream template, while the output table is the output result of executing complex

objects and has responsibility of publishing out the template executing results for downstream template to define data reference. Input and output table has the same parameter type as TDE parameter table.

(3)User interface: this is the interface of packed template data and outer data. When packing the template, users can design user interface in which the parameter information of the template and related auxiliary notes are displayed clearly, from which the related parameters of the template are revised easily in IDE environment and with which the specific operations of the template are implemented successfully.

(4)Operations: this is used to implement the process of template data. It is composed of the type operation units such as some formula operations, file analysis, database operation, command execution, CAD operation and report generation. It supports the sequential execution mode and branch execution mode.

Assembled Template Development Environment

Integrated into CAD environment, assembled template environment (also called Integrated Design Environment) is unit-template-based multi-template unified connected multi-discipline design, modeling, analysis and optimization integrated environment. By calling design, modeling, analysis templates and basic database, assembled template environment can rapidly set up design and analysis unified connected model, speed up the modification efficiency of design proposal greatly, conduct the backstage calculation of analysis software with work management system and carry on multi-discipline optimization by calling MDO software.

Unit templates constitute complex assembled templates, tool bags and group tool bags with data flow and control flow. Data flow defines the transitive relation among template instantiations, Pictured at Fig.2 (left). The blue line indicates that data transitive relation “exists” between template instantiation and instantiation. Users can define that relation, as shown in Fig.2 (middle).

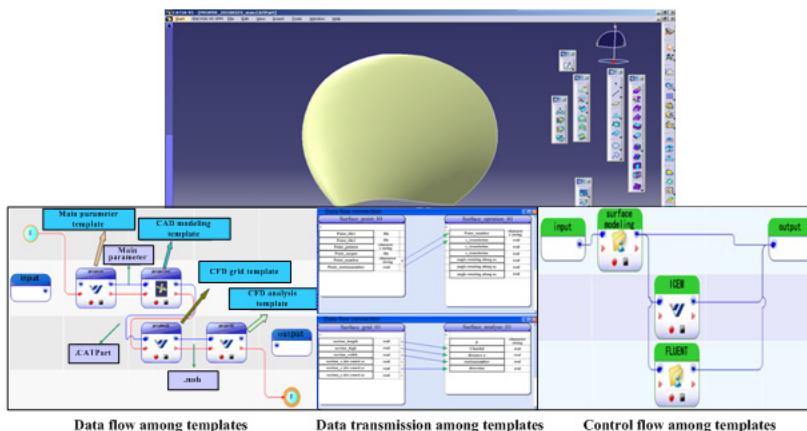


Fig. 2. Assembled template environment

Control flow indicates the execution sequence of template instantiations, as shown at Fig. 2 (right). Users can both define control flow connecting lines and check the execution sequence. Control flow can define starting point (icon “s”), ending point (icon “e”), parallel and estimate actuating logics such as branch and circulation.

Integrated Management System

Integrated management system is a synthetical management system of project management and process management, including functions like process/project definition, process driver and process monitoring/project analysis. Integrated management system can boost the execution efficiency and the standardization of work flow substantially. Supervisor of engineering of all levels can have a better trace of project status, task progress and resource utilization with comprehensive monitoring of the engineering project in real time, thus management efficiency of engineering development can be improved. It is briefly introduced in this paper.

Template Management System

Template management system is the system platform for template management, with which templates can be uploaded, managed, searched and downloaded. Templates are modular components packed by repeatable design, modeling and analysis operation in the design process. Utilizing the template in the integrated design environment can achieve rapid modeling of modular, rapid modification design proposal and calling of professional software to complete computational analysis. Template base mainly includes hull form design template, fluid analysis template, structural arrangement template, finite-element analysis template and multi-disciplinary optimization template.

Engineering Data Management System

Engineering data management system manages unified connected model and related data during the design process. Unified connected model can be called by engineering data management system, at the same time, engineering data management system manages all process data and final result according to unified connected model's structural relation with edition method. Moreover, data integration with PDM is also an important content of data management. Engineering data management system has interaction relationship with product structure management, document management and user management of PDM and the data of PDM can be called by standard service.

4 Application

In order to test the rationality, availability and reliability of the system, it takes a ship as example, sets up template base of hull generation, performance analysis, rapid cabin subdivision, rapid structural design, structural analysis and configuration optimization and builds template process control system. Integrated tools and developed template bases are shown in Table 1 and Table 2 as following. Main process and interface of the system are shown in Fig. 3.

Table 1. Integrated Tools

No.	Software	Description
1	CATIA	Used for ship hull geometry modeling and can be imported by FORAN system for follow-up design after calculation analysis confirmation.
2	ICEM	Used for meshing and pre-process.
3	FLUENT	Used for hydrodynamic computation.
4	ANSYS Self- compiled program	Used for meshing and pre-process/local feature analysis.
5	ISIGHT	Developed by our side.
6		Used for plan optimization design.

Table 2. Template databases

No.	Template databases	Description
1	Basic parameter design	Basic parameter input, mainly including blade form parameter(blade number, diameter, boss- diameter ratio, disk ratio, angle of inclination, etc.),and working condition parameter(rotating speed, ship velocity, flow fraction, etc.).
2	Table of offsets design	Table of offsets fast designing and solving, and outputting table of offsets.
3	Performance analysis	Performance optimization and analysis.
3.1	Hydrodynamics calculation of theory method	Performance calculation by lifting line, lifting surface and surface element.
3.2	CFD method	3D surface creation, CFD grid, flow field calculation and post- process based on commercial software.
3.3	Propeller performance result output	Creating performance result table according to theory method and CFD method.
4	Structure analysis	Structure design and analysis.
4.1	Strength checking of theory method	Strength checking and analysis by cantilever beam.
4.2	CAE FEM analysis	3D surface creation, FEM grid, loading boundary definition, Stress calculationand post- process based on commercial software.
4.3	Propeller strength result output	Creating strength result table according to theory method and FEM method.
5	Optimization strategy	Optimization strategy setup based on iSIGHT which is integrated in SYSWAKE.

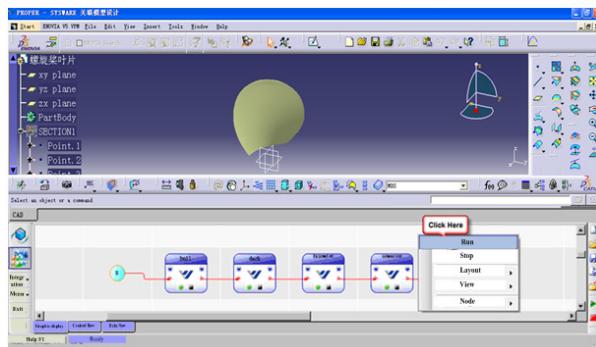


Fig. 3. Main interface of the System

5 Conclusion

Ship propeller integrated design system based on unified related model is put forward aimed at ship propeller design with the purpose of deducing ship propeller development coast, cutting down ship propeller development circle and boost ship propeller general performance. This paper shows a propeller unified related model and the implementation method—template technology, puts forward the frame structure and function modules of the ship propeller integrated design platform. With one example ship propeller designed on the platform, we conducted the preliminary analysis of general performance, realized the integration of tools such as CATIA, FLUENT, ANSYS etc., built parametric hull form surface and ship structural model, set the data connections among CAD, CAE and CFD and formed the template database of the fixation and accumulation of the knowledge and experience.

References

1. Huang, J.-F., He, G.: Technology of Naval Vessel Digital Ship and Study on its Platform Architecture. China, Ship Engineering 31, 41–42 (2009)
2. Shi, H., Liu, X.: Application Framework of Template Technique in CAD Field. China, Science & Technology Review 26, 89–91 (2008)
3. Simpson, R.G., Williams, A.A.: Application of Computational Fluid Dynamics to the Design of Pico Propeller Turbines. USA, Nottingham Trent University 31, 56–57 (2006)
4. Jayasumana, A.: Reorder density and reorder buffer-occupancy density metrics for packet reordering measurements. Internet Draft (2006)
5. Rhee, S.H., Joshi, S.: CFD validation for a marine propeller using an unstructured mesh based RANS method. In: Proceedings of Fluids Engineering Division Summer Meeting, FEDSM (2003)
6. Cavitation Tunnel Tests with Design Propeller for the MARIC 220K LNGC.R. SSPA, Report No.RE2008 5009-04-00-b, No.RE2008 5009-05-00-b (2010)
7. Towing Tank Tests with Design Propeller P2551 for the MARIC 220K LNGC.R.SSPA, Report No.RE2008 5009-02-00-a, No.RE2008 5009-02-00-a (2010)

8. Chen, B., Stern, F.: Computational Fluid Dynamics of Four-Quadrant Marine-Propulsor Flow. *Ship Research*, 218–228 (1999)
9. Ea, L., Hoekstra, M.: On the influence of the iterative error in the numerical uncertainty of ship viscous flow calculation, pp. 89–90 (2006)
10. Gert, K.: Blade Section Design of Marine Propellers with Maximum Cavitation Inception Speed. *Journal of Hydrodynamics* 1, 57–59 (2012)
11. Rhee, S.H., Joshi, S.C.: Fd validation for a marine propeller using an unstructured mesh based RANS method, pp. 35–37 (2003)

Skeleton Extraction of a Specified Object in the Gray Image Based on Geometric Features

Zhihui Yang, Fangfang Guo, and Ping Dong

Institute of Image Processing & Pattern Recognition,

North China University of Technology, Beijing

zhyang@ncut.edu.cn, guofangfang3@163.com

Abstract. In this paper, we present a two-stage method to extract the accurate and smooth skeleton of a specified object in the gray image based on geometric characteristics of the boundary. In the first stage, according to the statistical intensity disparity between the sample points and the object, an energy function is constructed, and then a novel segmentation model is proposed to extract any specified objects in the gray image by the variational method. In the second stage, on the basis of the segmentation, an improved skeleton extraction algorithm is given in virtue of the shortest path connection approach. Some examples show the robustness and insensitivity of the presented algorithm to the perturbation and noise, respectively.

Keywords: Image Segmentation, PDE, Skeleton.

1 Introduction

The concept of the skeleton was firstly proposed by Blum in 1967, which is defined as the locus of centers of maximal discs contained within the shape [2, 3]. The related investigations are important for the object shape description and widely applied in computer vision.

The mainstream of the research focuses on skeleton extraction algorithms [9-13] based on the study of geometric features of the skeleton [4-6]. Generally speaking, the skeleton is divided into two types: structural skeleton and texture skeleton. The former reflects the topology of the shape very well and is commonly used in shape analysis [9]. Moreover, the structure skeleton is usually explored in computer graphics, which is seldom mentioned in image processing.

Nowadays, some efforts have been put into the research of skeleton extraction in the gray image. Jang presented an extraction method without segmentation by extracting the ridge of the pseudo-distance map constructed in the image where the intensity of the object is always higher or lower than that of the background. the skeleton in this kind of images is traced from the skeleton strength map based on the gradient vector diffusion.

In our paper, we aim at finding a computational simple structural skeleton of the object specified in the image based on geometric features of the object's boundary. First inspired by [1], the specified object is obtained by an improved PDE

segmentation model, and then the skeleton of the specified object is formed by connecting endpoints through the shortest path algorithm, which can ensure the accuracy and smoothness of the skeleton. Here since we only consider the specified objects in images, it has much less computational complexity.

The remainder of this paper is organized as follows. In Section 2 we will introduce our segmentation proceeding of the specified object in the image. Section 3 presents a new skeleton extraction algorithm to find the specified object skeleton. Then experimental results are given in Section 4 to show the efficiency of our algorithm. Finally conclusions are drawn in Section 5.

2 Local Segmentation Model of a Specified Object

Before we get the skeleton of a specified object, we need to find the boundary of the object. To demonstrate the intrinsic principle, here we take the piecewise-constant image $I(x, y)$ for example.

First we draw an initial closed curve in the object in virtue of some prior information (Fig. 1), and then the region enclosed by the curve will expand according to the consistency of the object intensity until the curve reaches the boundary of the specified object. On the other hand, the intensity will change obviously once the curve passes the boundary. Hence, the problem can be summarized to find the region with the maximum area under the condition that the intensity variance in the region remains the same as that in the sub-region enclosed by the initial contour. The details of the segmentation model are given as follows.

Assume that I_0 and σ_0 are the mean and deviation of the intensity in the sub-region Ω_0 enclosed by the initial curve, respectively. Then the variance of the intensity inside the region Ω_+ enclosed by the evolving curve (Fig. 1) can be represented by

$$\sigma^2 = \frac{\int_{\Omega_+} (I(x, y) - I_0)^2 dx dy}{\int_{\Omega_+} dx dy}.$$

Since

$$|\sigma^2 - \sigma_0^2| = \frac{\left| \int_{\Omega_+} \left((I(x, y) - I_0)^2 - \sigma_0^2 \right) dx dy \right|}{\int_{\Omega_+} dx dy},$$

then

$$|\sigma^2 - \sigma_0^2| = 0 \iff \int_{\Omega_+} \left| (I(x, y) - I_0)^2 - \sigma_0^2 \right| dx dy = 0.$$

Here $|\sigma^2 - \sigma_0^2|$ measures the disparity between the intensity variance of pixels in Ω_+ and Ω_0 , and $|\sigma^2 - \sigma_0^2| = 0$ implies that they have the same intensity variance.

By the Lagrange multiplier method, we obtain the following energy function

$$E = \lambda \int_{\Omega_-} dx dy + \mu \int_{\Omega_+} \sqrt{|(I(x, y) - I_0)^2 - \sigma_0^2|} dx dy + \nu \int_{\partial\Omega_+} ds \quad (1)$$

where Ω_- and $\partial\Omega_+$ represent the complementary and the boundary of the domain Ω_+ respectively, and the third term aims to control the smoothing of the boundary. We may rewrite (1) in the form of the level set:

$$\begin{aligned} E &= \lambda \int_{\Omega} (1 - H(\phi)) dx dy + \mu \int_{\Omega} H(\phi) \sqrt{|(I(x, y) - I_0)^2 - \sigma_0^2|} dx dy \\ &\quad + \nu \int_{\Omega} |\nabla(H(\phi))| dx dy \\ &= \lambda \int_{\Omega} (1 - H(\phi)) dx dy + \mu \int_{\Omega} H(\phi) \sqrt{|(I(x, y) - I_0)^2 - \sigma_0^2|} dx dy \\ &\quad + \nu \int_{\Omega} \delta(\phi) |\nabla\phi| dx dy , \end{aligned}$$

where $H(x)$ and $\delta(x)$ are heavy side function and delta function respectively, i.e.,

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{otherwise} \end{cases}, \quad \delta(x) = H'(x). \quad \text{Define}$$

$$L(\phi, x, y) = \lambda(1 - H(\phi)) + \mu H(\phi) \sqrt{|(I(x, y) - I_0)^2 - \sigma_0^2|} + \nu \delta(\phi) |\nabla\phi|,$$

Then the segmentation problem can be solved by the following variatioal Model:

$$\arg \min_{\phi} L . \quad (2)$$

Consider the Euler-Lagrange equation of (2) and obtain

$$\begin{aligned} \phi_t &= -\frac{\partial L}{\partial \phi} + \frac{\partial}{\partial x} \frac{\partial L}{\partial \phi_x} + \frac{\partial}{\partial y} \frac{\partial L}{\partial \phi_y} \\ &= \lambda \delta(\phi) - \mu \delta(\phi) \sqrt{|(I - I_0)^2 - \sigma_0^2|} - \nu \delta'(\phi) |\nabla\phi| \\ &\quad + \nu \frac{\partial}{\partial x} \left(\delta(\phi) \frac{\phi_x}{|\nabla\phi|} \right) + \nu \frac{\partial}{\partial y} \left(\delta(\phi) \frac{\phi_y}{|\nabla\phi|} \right) \\ &= \lambda \delta(\phi) - \mu \delta(\phi) \sqrt{|(I - I_0)^2 - \sigma_0^2|} - \nu \delta'(\phi) |\nabla\phi| \\ &\quad + \nu \delta'(\phi) |\nabla\phi| + \nu \delta(\phi) \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) \\ &= \lambda \delta(\phi) - \mu \delta(\phi) \sqrt{|(I - I_0)^2 - \sigma_0^2|} + \nu \delta(\phi) \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) \end{aligned} \quad (3)$$

with the initial condition $\phi(0, x, y) = \phi_0(x, y)$, where $\phi_0(x, y) = 0$ represents the initial contour, i.e. the boundary of the sub-domain of the segmented object. Here, we define ϕ_0 by the sign distance function. Fig. 1 shows the result applying (3) to the

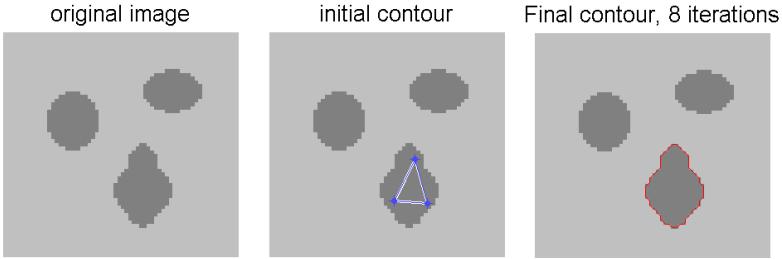


Fig. 1. Local segmentation in a piecewise-constant image

piecewise constant image. It can be seen that we only obtain the object we marked, not all objects in the image, which differs from previous methods.

However, when applying the equation (3) to the general gray image, we found that the boundary of the sub-region, that is the initial contour, is not always move outward. It obviously violates the truth that the sub-region is part of the object being segmented. After analysis and experiments, we replaced λ by $\lambda \cdot p(i, j)$, where

$$p(i, j) = \begin{cases} \frac{p_1(k)}{p_2(k)} & I(i, j) = k \text{ and } \phi(i, j) > 0 \\ \max \frac{p_1(k)}{p_2(k)} & \phi(i, j) \leq 0 \end{cases},$$

$p_1(k)$ and $p_2(k)$ denote the frequencies of points with intensity k in the sub-image region and the whole image region respectively, and $p(i, j)$ indicates the possibility of the pixel in the object region segmented. It will ensure the curve to evolve outward (Fig. 2). Figure 2 not only demonstrates a good segmentation result for a general image, but also exhibits robustness of the algorithm to the noises.

3 Skeleton Extraction Based on the Local Segmentation

3.1 Initial Skeleton Extraction

After the extraction of the boundary, we realize the distance transform by solving PDE $|\nabla\phi| = 1$ with the fast sweeping method [6], where ϕ is the signed distance function to the boundary. In practice, the skeleton is the set of non-differentiable points of the signed distance function ϕ , where the gradient changes greatly. We have provided some pictures to show the distance transform and grads module results in Figure 3. It is obvious that the gads module on the ridge of the distance function is much smaller than that on other points, so we consider $|\nabla\phi| < \alpha$ on the skeleton points, where α determined the domain of the skeleton. However, from the pictures we can also find there are some points with $|\nabla\phi| < 1$ near the edge, which is caused by the serrated boundary. But it will make no contribution to the structural skeleton, so we prune them based on the area criterion since the connective domains formed by them are much smaller than the initial skeleton.

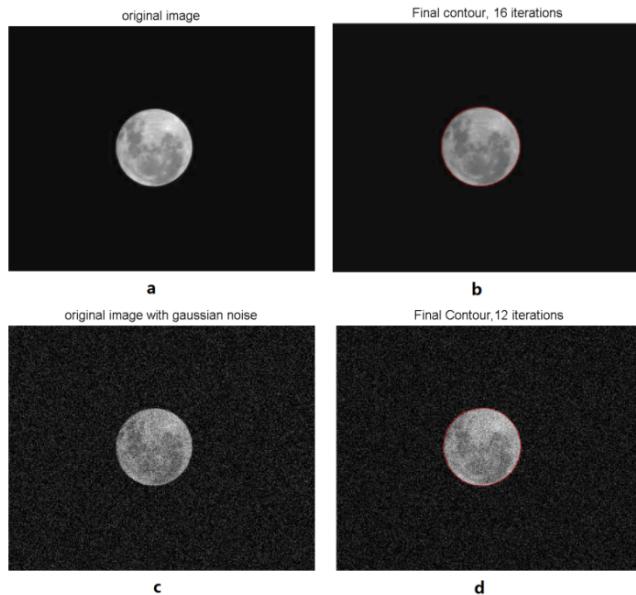


Fig. 2. Gray image segmentation: a) Original image b) Segmentation result under the improved segmentation model c) Original image with Gaussian noise d) Segmentation result for the noisy image

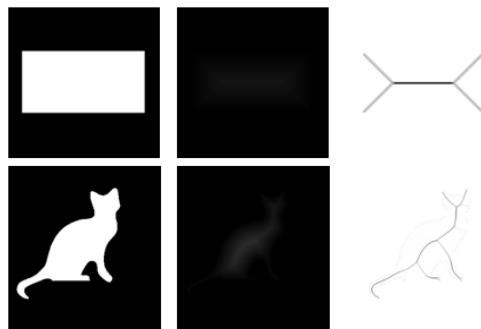


Fig. 3. Signed distance function and its grads module: original images in the first column, the signed distance functions in the second, and the grads module of the SDF in the last one

3.2 The Searching of Endpoints Based on the Harris Corner Detection Algorithm

The searching of an endpoint on the skeleton is similar to that of the vertex with maximum of curvature in the boundary [6]. Notice that the vertex is on the boundary, thus we may use the Harris corner detection algorithm to find the vertex [8].

The basic idea of the Harris corner detection is to construct an autocorrelation analytic function by considering changes along a particular direction in the image. The intensity changing along the direction can be approximated by

$$E_{u,v} = \sum_{i=-\omega}^{\omega} \sum_{j=-\omega}^{\omega} \left(\frac{\partial f(x+i, y+j)}{\partial x} u + \frac{\partial f(x+i, y+j)}{\partial y} v \right)^2 \quad (4)$$

i.e. $E_{u,v} = A(x,y)u^2 + 2C(x,y)uv + B(x,y)v^2$, where

$$A(x,y) = \sum_{i=-\omega}^{\omega} \sum_{j=-\omega}^{\omega} \left(\frac{\partial f(x+i, y+j)}{\partial x} \right)^2, \quad B(x,y) = \sum_{i=-\omega}^{\omega} \sum_{j=-\omega}^{\omega} \left(\frac{\partial f(x+i, y+j)}{\partial y} \right)^2.$$

$$C(x,y) = \sum_{i=-\omega}^{\omega} \sum_{j=-\omega}^{\omega} \frac{\partial f(x+i, y+j)}{\partial x} \cdot \frac{\partial f(x+i, y+j)}{\partial y}.$$

where (u, v) is the direction, f is the gray image, and ω determines the size of the local neighborhood. The minimum of E is a measure of the curvature since it is large when the point is a corner and small otherwise.

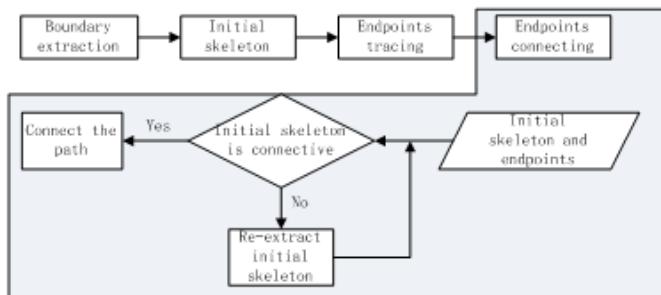
It can be observed that (4) is a quadratic function with two eigenvalues corresponding to two principal directions. The minimal eigenvalue is identical with the minimum of E that is, the curvature. The direction represents the tangent direction while the other is the normal direction for the orthogonality of the vectors. We search the endpoints along the normal direction until the local maximum of the grads module is reached.

3.2 Skeleton Extraction Algorithm

Given the skeleton endpoints x_1, x_2, \dots, x_n , we next find the skeleton through the shortest path method by regarding the image as a graph that every pixel is its node. The length of the edge from p_i to its adjacent point p_j , 8-connected point in the

image, is defined as $d(p_i, p_j) = \begin{cases} |\nabla \phi| & p_i, p_j \in S \\ +\infty & \text{otherwise} \end{cases}$. Then we will use the Dijkstra's

shortest path algorithm to acquire the skeleton. The difficulty in the procedure is that the initial skeleton may be not connective for the small parameter α in the step of extracting the initial skeleton, which will lead to the path impassable. In this case, we have to increase the scale until the initial skeleton satisfies the requirement. We summarize our skeletonization algorithm in the following diagram:



4 Skeleton Extraction Experiments and Results

In Figure 4, we provide a comparative result between our method and the classical morphological method for binary images. From Fig. 4 we can see that the morphological method performs vast redundancy stems caused by the small serrated edge, while ours is not affected by the boundary and is smooth enough. In addition, our algorithm is also suitable to the gray images with noise, as is shown in Figure 5. A smooth and connective structural skeleton of the object without additional pruning steps is obtained. Furthermore, skeletons of the toy in the original image and noisy image are similar to each other. It demonstrates that the noise has not an obvious effect on the extraction of the skeleton algorithm.



Fig. 4. Skeleton extraction with our method and morphological method: original image, the skeleton extracted by our method in the second column, and the skeleton extracted by the morphological method in the last one.

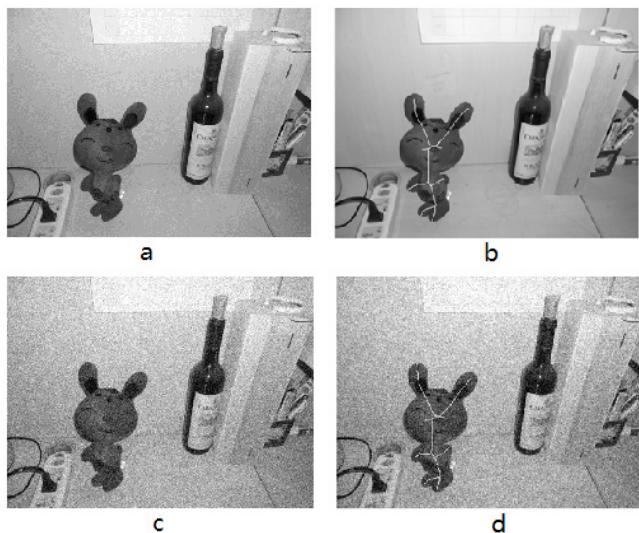


Fig. 5. Skeleton of the object in the gray image: a) the original image; b) Toy skeleton extraction result; c) Original image with Gaussian noise. d) Skeleton extraction result in the noisy image.

5 Conclusion

This paper proposes a novel method to extract the structural skeleton in gray images. Two steps have been taken in our extraction algorithm. In the first step of the segmentation, the interested object can be singled out by the improved segmentation model, which shows a good result insensitive to the noise. Then a robust skeleton extraction algorithm is proposed to find the connective and smooth skeleton by the shortest path approach. The result is more accurate and effective to avoid the pruning compared with the morphological method for the gray images.

Acknowledgments. This work was supported by National Natural Science Foundation No. 61170327, PHR(IHLB) (No. 201008199) and No. 2010D005002000008.

References

1. Chan, T.F., Vese, L.A.: Active Contour without Edges. *IEEE Transactions on Image Processing* 10, 266–277 (2001)
2. Blum, H.: A Transformation for Extracting New Descriptions of Shape, Models for Perception of Speech and Visual Form. In: Wathen-Dunn, W. (ed.), pp. 362–380. MIT Press (1967)
3. Blum, H.: Biological Shape and Visual Science. *J. Theor. Biol.* 38, 205–287 (1973)
4. Torsello, A., Hancock, E.R.: A Skeletal Measure of 2D Shape Similarity. *Computer Vision and Image Understanding* 95, 1–29 (2004)
5. Kuijper, A., Olsen, O.F., Giblin, P., Nielsen, M.: Alternative 2D Shape Representations using the Symmetry Set. *J. Math. Imaging Vis.* 26, 127–147 (2006)
6. Giblin, P.J., Kimia, B.B.: On the Local Form and Transitions of Symmetry Sets. Medial Axes, and Shocks. *International Journal of Computer Vision* 54, 143–157 (2003)
7. Zhao, H.K.: A Fast Sweeping Method for Eikonal equations. *Mathematics of Computation* 74, 603–627 (2004)
8. Nixon, M., Aguado, A.: Feature Extraction and Image Processing. Academic Press, London (2008)
9. Goh, W.-B., Chan, K.-Y.: Structural and Textural Skeletons for Noisy Shapes. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 454–461. Springer, Heidelberg (2005)
10. Goh, W., Chan, K.: The Multiresolution Gradient Vector Field Skeleton. *Pattern Recognition* 40, 1255–1269 (2007)
11. Klette, G.: Skeletons in Digital Image Processing. Technical Report CITR-TR-112, Centre for Image Technology and Robotics of The University of Auckland (2002)
12. Serra, J.: Image Analysis and Mathematical Morphology. Academic Press, Orlando (1983)
13. Siddiqi, K., Bouix, S., Tannenbaum, A., Zucker, S.W.: The Hamilton-Jacobi skeleton. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision 2, pp. 828–834 (1999)

Detection and Correction Scheme of Internet Chat Lingo Based on Statistic and Pinyin Similarity

Bo Han and Zhengwen Li

International School of Software, Wuhan University, Wuhan, China
bhan@whu.edu.cn

Abstract. The development of Internet promotes the usage of Internet chat lingo. This type of language is diversified and irregular for natural language processing. In this paper, according to the characteristics of the Chinese and Internet chat lingo, we proposed a method for lingo detection and correction based on the statistic and pinyin similarity. This method applied Bigram model to detect the boundary of lingos, and then corrected them by using pinyin similarity. According to the experimental results and analysis, our method can effectively detect and correct chat lingos.

Keywords: Internet chat lingo, N-gram model, Pinyin Similarity, Levenshtein Distance.

1 Introduction

In the recent years, with the development of Internet and more and more people use it, chat lingos rise as new terms in our daily work and life. They are rich in BBS and forums. These words are generally divided into symbolic term, homophonic words and old words with a new meaning. The homophonic words is the most common used. For example, “神马” represents “什么”, “斑竹” indicates “版主”.

Since Internet has boomed to be the main source of information, scientists are paying more and more attention to Internet language understanding and correction. However, Internet chat lingo is diversified and irregular, difficult for natural language processing.

In the research area about internet language, many scientists have done much work on analyzing characteristic, origin, culture background of Internet language[1-2]. However, the research on how to use text mining techniques to detect and correct Internet language is not abundant. Current most text correction algorithms are applied on formal text [3-7]. In this paper, according to the characteristics of the Chinese language and Internet chat lingos, we proposed a method for detection and correction of chat lingos based on the statistic and pinyin similarity. This method applied Bigram model to detect the boundary of chat lingos in a sentence, and then corrected them by computing pinyin similarity.

2 Methodology for Detection and Correction of Chat Lingos

We first explain the key points of our method and then list the main procedure for our algorithm.

2.1 Detection the Boundary of Chat Lingos

N-gram linguistics model is the most often used statistic model in natural language processing for boundary detection [8-9]. Its core concept is conditional probability model which assumes that the i^{th} word is dependent on the $n-1$ words forward. As a result, this model can describe the context of the sentence precisely. Providing the variability of phrase composition, the computation for n-gram model will be very huge. By the rule of the thumb, n is set to 2 or 3 (namely Bigram model or Trigram model).

The basic unit in Chinese is word. Observing 2 million papers of People Daily published in 1998, we find Chinese language often use two-word combination for nouns and verbs. The language uses a single word as an adverb, particle, preposition, conjunction. The length of scattered text, which is composed by independent single words, is often 3 and it is no more than 5. From the research and statistic of Professor Liu Yun [10], Modern Chinese grammar information dictionary has included 51,696 words in total, with 3,803 single words (accounting for 7.3%), 32,711 double words, 7,926 three words, 7,220 four words. After we applied word-segmentation program in text, we find that most of the words and expressions are single word. Based on the characteristic that scattered text and chat lingos are often segmented to single word, we define a method to probe the boundary of lingos: Firstly, we segment the text. Secondly, we parse every two words (w_1, w_2) and judge whether they are the boundary of Internet chat lingos.

In details, we firstly count the total number that w_1 and w_2 occur continuously. If the number is 0, we need to judge the length and the appear times of w_2 . If w_2 's length is 1, and it appears less than 5 times, we set w_1 as the words of start boundary when the start boundary is not set, otherwise, we read a next pair of words. If the length of w_2 is big than 1, and it appears more than 3 times, we will set w_1 as the word of end boundary when the start-boundary is set. Otherwise, we read a next pair of words.

If the total number that w_1 and w_2 occur continuously is larger than 0, we will set w_1 as the words of end boundary when the start boundary is set, otherwise, we read next pair of words.

2.2 Calculate the Similarity of Two Pinyins

Compare the similarity of two pinyin is similar to compare the similarity of two strings. We can determine the degree of the similarity of two pinyin by calculate their Levenshtein Distance [11-12].

Levenshtein Distance is one of the edit-distance functions in which distance is the cost of the best sequence of edit operations that convert one string to another string.

These edit operations are character insertion, deletion, and substitution, and each operation is assigned a cost. We set the cost of each operation to one.

For example, suppose that we have two strings: s and t, s is assigned to “shenme”, and t is assigned to “shenma”. s can be converted to t by replacing the character ‘e’ with character ‘a’. However there are other ways to achieve this goal. We can delete the last character ‘e’ in s, and then add character ‘a’ to the end of s. we define the cost of the edit operation as minimum edit distance.

In general, dynamic programming is a method for calculating the Levenshtein Distance.

3 Select the Correction Candidate Words

The selection of the correction candidate words is divided into three steps,

Compare the similarity between chat lingos and the words in a self-organized dictionary. If a matched word is retrieved, we return the words as the correction candidate word.

If the first step does not work, we compare the similarity between lingos and the words in the follow-words list of the words in start boundary. If a matched word is retrieved, we return the word as the correction candidate word. Here is an example, for a sentence, “这是神马东西呢”，it will be segmented to “这/是/神/马/东西/呢”. After applying the detection algorithm onto it, we can get the “是” as the words in the start boundary, and “东西” as the words in the end boundary. we will compare the similarity of the pinyin of “神马” and the pinyin of the word in the follow-words list of “是”. In the Bigram model we find that “什么” exists in the corpus following “是”. We pick “什么” as the correction candidate word to return since the pinyin of “什么” is similar to the pinyin of “神马”.

If both the first and second step do not work, we pick the word which occurs more often than any other words in the intersection of the follow-words list of the word in the start boundary and the pre-order-words list of the words in the end boundary as the correction candidate word to return.

3.1 The Main Procedure of Algorithm

The main procedure of our algorithm is listed as below,

Step1: Pre-process training documents to generate Bigram Model and import self-organized dictionary.

Step2: Split input document by period mark and send each piece to variable str

Step3: Divide str to pieces by '，' '；' '。' '？' '！'， and send each piece to variable temp

Step4: Segment temp to words

Step5: Based on the proposed chat lingo detection method, we extract chat lingo's pinyin and store it in variable error_PY

Step6: Compare the similarity between error_PY and words in self-organized dictionary, if matched, we return the corresponding text as candidate, and jump to step 9, otherwise, we jump to step 7

Step7: Compare the similarity between error_PY and follow-words list of the start boundary words. If matched, we select them as candidate words and jump to step 9, otherwise, we jump to step8

Step8: Get the follow-words list of the start boundary word—NEXT and prefix-words list PRE of last boundary word, sort NEXT and PRE in descending order and search the first same word in both NEXT and PRE as candidate word and return it. If no common word is found, empty string will be returned.

Step9: If candidate word is not empty, we replace chat lingos with candidate words

Step10: Return updated sentence and read next temp

4 Experimental Results

In this experiment, we have chosen about 200 thousand words for training to build the Bigram Model. These words were chosen from People Daily in 1998, and they were processed by teachers in Peking University. We exacted the pinyin of a word with a tool called pinyin4j, we modify the tool to facilitate the experiment. The Threshold of the Levenshtein Distance was set to 3, which means that we define two pinyin is similar when the Levenshtein Distance is less than 3. We segment the sentence by software ICTCLAS developed by the China Academy of Science. And there are five records in the self-organized dictionary:

雅蠛蝶&不要啊&yamiedie^{4j}
 楼主&帖子的主人&louzhu^{4j}
 版主&论坛的主人&banzhu^{4j}
 碟堡了&太厉害了&diaobaole^{4j}
 酱紫&这样子&jiangzi^{4j}

Experiment 1: 这是神马东西呢？

After word segmentation: 这/ 是/ 神/ 马/ 东西/ 呢

Detect the boundary: we will find that “这” and “是” appear together for 45 times. So they should not be setted as the boundary. Then we read next pair of words---“是” and “神”, we find that they appear together for 0 times, and the length of “神” is 1, its total times appear in the corpus is 0, which is small than 5, so we set “是” as the words of the start boundary. Then we read next pair of words---“神” and “马”, the number of times that they appear together is 0, so we read the next pair of words---“马” and “东西”, the number of times that they appear together is also 0, but the length of “东西” is more than one, and its total times that appear in the corpus is 5, which is more than 3, so we set the “东西” as the end of the boundary.

Correct the chat ling: first , we compare the similarity of the pinyin of Internet chat lingos and the words in the self—organized dictionary. Specifically, we compare the

Levenshtein Distance between the pinyin of “神马”, which is shenma with the words in the dictionary. The result we will get are 7,6,6,8,6. All the edit-distance is more than 3, so in next step , we compare the Levenshtein Distance between the pinyin of “神马” and the pinyin of words in the follow-words list of “东西”. We find that the edit-distance between “神马” and “什么”, which is one of the follow-words of “是”, is 1, so we determine “什么” as the correction candidate word.

Experiment 2: 要是斑竹觉得可以的话,那就去做吧

After word segmentation: 要/ 是/ 斑竹/ 觉得/ 可以/ 的/ 话/

Detect the boundary: do as the first experiment, we will find that the start boundary word is “是”,and the end boundary word is “觉得”.

Correct chat lingos: when we compare the pinyin similarity between “斑竹” and the word in the self-organized dictionary, we find that the edit-distance between “斑竹” and “版主” is 0, so we can determine “论坛的主人” as the correction candidate word.

From the above experiments, we can conduct that the key of the method to success is the scale and content of the corpus. Since we selected a small-scale corpus to calculate the frequency of words and compute the adjacency relationship between terms, so we can not get a full reflection of all Chinese sentence. For instance, “这种大尺度而且很容易引起争论的问题果然引起了很多人的围观”, this sentence will be convert to “这种大尺度而且很容易引起争论的问题果然引起了很多人的围观”.

5 Conclusion

In this paper, we proposed a method for detection and correction of Internet chat lingos based on statistics and pinyin similarity. Our experimental results show the effectiveness of our approach. However, the method is not perfect. For example, the method now is just on the base of lexical analysis, not yet into the level of semantic analysis. And the size of the corpus is still small, so we can not get a fully reflection of sentence context. The self-organized dictionary is also going to be improved. The function for detecting the boundary seems too simple. These are the research points where we will improve the method. In conclusion, the method proposed in this article is just launching the research of correction Internet chat lingos, we will continue in-depth analysis and expect to achieve better results.

Acknowledgment. This work was supported by grant 2009CB723905 from National Basic Research Program of China (973 Program), by grant 2011CDB447 from natural science foundation of Hubei province, by grant 216271207 for talent development at Wuhan University.

References

1. Xie, Y.: Research for Internet Language. *Medical Information* 20(5) (2007) (Chinese)
2. Wu, Z.H.: The Study of Phonetic Metaphors in Cyber Language. *Journal of Henan Institute of Engineering (Social Science Edition)* 25(1) (2010) (Chinese)
3. Chen, Z.P., Lv, Y.Q., Liu, H.S., et al.: Chinese Spelling Correction in Search Engines Based on N-gram Model. *Journal of CAEIT* 4(3) (2009) (Chinese)
4. Ma, J.S., Zhang, Y., Liu, T., et al.: Detecting Chinese Text Errors Based on Trigram and Dependency Parsing. *Journal of the China Society for Scientific and Technical Information* 23(6) (2004) (Chinese)
5. Chen, T.Y., Chen, R., Pan, L.L., et al.: Archaic Chinese Punctuating Sentences Based on Context N-gram Model. *Computer Engineering* 33(3) (2007) (Chinese)
6. Zhang, Y.S., Cao, D.Y., Yu, S.W.: A Hybrid Model of Combining Rule-based and Statistics-based Approaches for Automatic Detecting Errors in Chinese Text. *Journal of Chinese Information Processing* 20(4) (2006) (Chinese)
7. Zhang, Y.S., Ding, B.Q.: Automatic Errors Detecting of Chinese Texts Based on the Bi-neighborship. *Journal of Chinese Information Processing* 15(3) (2000) (Chinese)
8. Feng, J.H., Gulila, A., Mayra, H.K.: Organization Name Recognition based on N-gram Model. *Computer Engineering and Application* 46(31), 135–138 (2010)
9. Diane, M.N., Amanda, S.: TechWriter: An Evolving System for Writing Assistance for Advanced Learners of English. *CALICO Journal* 26(3), 611–625 (2009)
10. Liu, Y., Yu, S.W., Zhu, X.F.: Construction of the Contemporary Chinese Compound Words Database (Chinese)
11. Zhou, H.P.: Study on Application of Levenshtein Distance in Programming Test Automatic Scoring. *Computer Applications and Software* 28(5) (2011) (Chinese)
12. Mayire, Y., Mijiti, A., Askar, H.: A Minimum Edit Distance Based Uighur Spelling Check. *Journal of Chinese Information Processing* 22(3) (2008) (Chinese)

A Filling Algorithm of Mining Constraint Frequent Neighboring Class Set

Ling Hu¹, Yao Li^{1,*}, Jiang Xiong², and Gang Fang²

¹ College of Computer Science, Neijiang Normal University, China

² College of Mathematics and Computer Science, Chongqing Three Gorges University, China
{hhuling,cqwzjsjfg}@163.com, lycn_99@126.com,
xjxj123@ms.mmit.stc.sh.cn

Abstract. In mining constraint frequent neighboring class set, it has been reported that the present algorithm is unsuitable for mining long constraint frequent neighboring class set in spatial database. This paper proposes a filling algorithm of mining constraint frequent neighboring class set, which is suitable for mining any length constraint frequent neighboring class set. In order to improve efficiency of mining algorithm, the algorithm uses computing subset of neighboring class set to generate candidate via top-down strategy. Meanwhile, it uses the method of filling subset to double generate candidate frequent neighboring class set, and it also uses ‘and’ operation of corresponding integer to compute support. The result of experiment indicates that the algorithm is faster and more efficient than other algorithms when mining constraint frequent neighboring class sets.

Keywords: neighboring class set, constraint condition, filling subset, top-down, spatial data mining.

1 Introduction

The location mold of the spatial relationship is the implied rule of the GIS which reflect the spatial objection structure and the spatial relationship. It is the mainly represent of the hierarchy and the mutual relationship among the different subset of the spatial relationship or the spatial data [1, 2, 3]. At present, the common approaches to study the spatial related rule are layer covering method based on clustering [4], the mining method based on the spatial transaction [3, 5 and 6] and non-spatial transaction mining method [4]. The first two methods can be used to mine the frequent neighboring class set. The spatial related rules from the document [5, 6, 7 and 8] study the type relationships which are both close to the target entity and non-target entity, while the frequent neighboring class set study the relationship of the adjacent entities which related to their own types. Even the algorithm from the document [1, 3, 9, 10] can search the spatial frequent neighboring class set, it can hardly mine the restrained frequent neighboring class set. The MFNCSWCC

* Corresponding author.

algorithm which we propose in the document [11] can mine the restrained frequent neighboring class set. But it comes out many repetitions in computing and the redundant candidate. So we propose a filling algorithm of mining constraint frequent neighboring class set, FAMCFNCS.

2 The Description of the Relevant Issue and Knowledge

Spatial data sets are composed of spatial object entities appointed in the spatial scope. This paper describes the data structure of these spatial entities and defined as $\langle ID_{Instance}, ID_{Class} \text{ and Location} \rangle$. The ID_{Class} spatial data set used to distinguish different object categories in the spatial data set, denoted by C_i , $ID_{Instance}$ is used to distinguish different objects entities in the same spatial class, denoted as i_k , Location is used to describe the location coordinates of the object entities. A spatial object entity is called an instance of the space type which it corresponding to. So the spatial data set is composed of the instance ID_{Class} . The set of ID_{Class} is called spatial class set composition, if there are m different types of spatial objects, denoted as $C = \{C_1, C_2, \dots, C_m\}$.

Definition 1: Neighboring class set is the subset of the spatial class set in the spatial data set. Denoted as $\{C_{t1}, C_{t2}, \dots, C_{tk}\}$ the $t_k \leq m$, taken as NCS. We take $I = \{i_{t1}, i_{t2}, \dots, i_{tk}\}$ is a instance set of the neighboring class set $NCS = \{C_{t1}, C_{t2}, \dots, C_{tk}\}$, spatial data types concentrated space for a subset of C with $\{C_{t1}, C_{t2}, \dots, C_{tk}\}$ said, where $t_k \leq m$, denoted as NCS. Write $I = \{i_{11}, i_{12}, \dots, i_{1k}\}$ the neighboring class set $NCS = \{C_{11}, C_{12}, \dots, C_{1k}\}$, the i_{1j} in it is an instance of the spatial class set C_{1j} ($j \in 1, 2, \dots, k$).

For example if the $NCS = \{X, Y, Z\}$, then $I = \{X_2, Y_3, Z_4\}$ is an instance set of NCS.

Definition 2: The length of neighboring class set refers to the different ID_{Class} number of the neighboring class set; if the length of a neighboring class set is k , we can call it k -neighboring class set, denoted as k -NCS.

Definition 3: The correct instance sets of the neighboring class set refers to an instance set which belongs to the neighboring class set $I = \{i_{t1}, i_{t2}, \dots, i_{tk}\}$. If the $\forall i_p$ and $i_q (i_p, i_q \in I)$ both have the distance $(i_p, i_q) \leq d$ are tenable, we deem that I is the correct instance set of the neighboring class set NCS; here d is the threshold which measures the distance between the two adjacent instances appointed by the user. Distance (i_p, i_q) represents the Euclidean distance of the spatial instances i_p and i_q .

Definition 4: The constraint neighboring class set is non-void proper subset of neighboring class set. This space type is the distinguishing requirement proposed by the user. It is often called user-specified constraints, denoted as C -NCS.

Definition 5: The support number of the constraint neighboring class set is defined as: the number of the correct instances set which supports the constraint neighboring class set NCS, denoted as Support.

Definition 6: Constraint frequent neighboring class set refers to the support number greater than or equal to the number of user-specified minimum support number and also support the constraint neighboring class set.

Definition 7: The (L-1)-subset of the L-integer, L-integer refers to the binary number which contains L a "1"; the (L-1) subset is a subset which removes one "1" from L-integer binary

For example, $14 = (1110)_2$, 14 is the 3-integer, the 2 - integer there are three subsets: $12 = (1100)_2$, $10 = (1010)_2$, $6 = (0110)_2$.

If I_k is the instance set of k-NCS, then the following properties hold.

Property 1: If k-NCS is a frequent neighboring class set, then (k-1)-NCS is also a frequent neighboring class sets; here $(k-1)\text{-NCS} \subset k\text{-NCS}$.

Property 2: If (k-1)-NCS is a non-frequent neighboring class set, so the k-NCS is also the non-frequent neighboring class sets; here $(k-1)\text{-NCS} \subset k\text{-NCS}$.

Property 3: Only if each I_k ($\forall I_k \subset I_{k+1}$) is correct, in which $k \geq 2$, then the instance set $(k+1)$ -NCS is correct.

According to previous knowledge and problem description, constraints, set frequent problem of mining adjacent categories can be described as follows:

Given: (1) spatial class set $C = \{C_1, C_2, \dots, C_m\}$ and spatial instance set $I = \{i_1, i_2, \dots, i_n\}$, for each i_k ($i_k \in I$) denoted as the data structure $\langle ID_{Instance}, ID_{Class}, Location \rangle$. (2) the adjacent distance threshold of the spatial object. (3) constraint neighboring class set C-NCS. (4) the minimum supports the user give.

Solution: to find all the neighboring class set which its support is greater than or equal to s and supports the constraints neighboring class set.

3 The Filling Algorithm of the Restrained Frequent Neighboring Class Set

3.1 The Forming of Mining Database of the Neighboring Class Set

The algorithm uses the following data structure to save the neighboring class set:

Structure Neighboring Class Sets

```
{
int value; // to save the corresponding integer of the neighboring class set
int count; // to save the number of the correct examples of the neighboring class set
} NCS;
```

The forming process of the mining database of the neighboring class set:

Step1: Scanning all the correct instance set I_k , to the arbitrary i_j ($i_j \in I_k$), through its property ID_{Class} , we can get the type C_t which in this example refers to and the serial number of location Location ($Location \in [1, m]$) in the spatial class set C, the neighboring class set NCS_k what the instance set of the I_k support is composed by these type C_t .

Step2: Defining the Integer= $\sum_{t=1}^L 2^{Location_t - 1}$ which is the integer of the NCS_k.

Step3: If this integer is not saved in the NCS array, we save the integer in the NCS, then save the relevant information as following: the field value =Integer, count=1; Otherwise we add the one to the value=Integer.

Step4: Scanning the I_{k+1} , which is the next correct example, operating the Step1-Step3 repeatedly until we finish in scanning all the correct examples in the spatial data set, at last outputting the mining database NCS.

3.2 The Method of Calculating Neighboring Class Sets Support Number

When computing the support number and examine the restricted condition, we need to computer the two neighboring class sets to see whether these two sets can satisfy the support relationship. According to the corresponding integer of neighboring class set, Algorithm through logic “and” operation can compute the relationship between the two neighboring class sets quickly. Therefore, the efficiency which compute the support and examine the restricted condition will be improved. This algorithm is equal to the theory of set computing. The definition is as follows:

Suppose the neighboring class set NCS_p correspond the integer I_p , the integer which NCS_q correspond is I_q , if $I_p \wedge I_q = I_p$ and then the NCS_q support NCS_p

For example, suppose the spatial class set $C=\{V,W, X, Y, Z\}$, if the neighboring class set $NCS_p=\{V,W,X\}$, the corresponding integer $Integer_p=7$, the $NCS_q=\{V,W,X,Z\}$, its corresponding integer $Integer_q=23$, the $I_p \wedge I_q = I_p$, so the $7 \wedge 23=7$, and then the NCS_q support NCS_p , that is $\{V,W,X,Z\}$ support $\{V,W,X\}$.

3.3 The Process of Double-Direction Generating the Constraint Frequent Neighboring Class Set

When mining the constraint frequent neighboring class set using the algorithm MFNCSWCC [11], the method which this algorithm generates candidate frequent neighboring class set is similar to the algorithm Apriori. That is, the algorithm uses down-top searching strategy to generate candidate neighboring class set, generating the candidate frequent neighboring class set which includes the constraint condition $(k+1)$ - from the frequent neighboring class set through iteration. The length of the frequent neighboring class set is ascending, and also as the length added, the number and the repetition computing work of the candidate frequent neighboring class set will be increased and the efficiency of the algorithm will be restricted.

The method we put forward to generate the candidate frequent neighboring class set is: to acquire the candidate frequent neighboring class set duly in two-way through generating the subset of non-frequent neighboring class set and its filling method. That is, with the top-down searching strategy to compute the subset of the non-frequent neighboring class set to generate the candidate frequent neighboring class set, and fill the aforementioned subset to generate frequent neighboring class set. The procedures are as follows:

FNCS: to save the restrained frequent neighboring class set through two-way mining.

NFNCS_d: to save the non-frequent neighboring class set which generates via top-down approach.

NFNCS_f: to save the non-frequent neighboring class set through the subset filling.

Step1: Taking the spatial class set C as candidate frequent neighboring class set m-CNCS, computing its integer (m-CNCS), and computing the support. If this support is the frequent neighboring class set and then save into FNCS, and terminate the algorithm. Otherwise, we save it into NFNCS_d.

Step2: To the every item L-CNCS (the initial value of L is m), computing the subset (L-1)-CNCS of L-CNCS. If the (L-1)-CNCS isn't any one of the following condition, then it's necessary to compute its support number. Otherwise we have to calculate;

it does not support the constraint frequent neighboring class set C-NCS.

it is the subset of some items in FNCS.

it is the superset of the some items in NFNCS_f.

Step3: To fill the subset of the (L-1)-CNCS of the every L-CNCS (L-1)- and make it equal to the m-CNCS, to form all the class set to the candidate neighboring class set (m-L+1)-CNCS. If it is not the subset of some items in the FNCS, and then computes its support; if it is the frequent item, we can save it in the FNCS, or else we save it in the NFNCS_f.

Step4: Conducting the Step2-3 repeatedly until L= [m/2]. The two-way mining the constraint frequent neighboring class set is finished.

Step5: Output the FNCS.

4 The Analysis of the Algorithm Performance and the Comparison of the Experiment

The literature to study the constraint frequent neighboring class set is that, we make analysis to the algorithm FAMCFNCS we put forward in this dissertation. Then we make a simulated experiment between the algorithm FAMCFNCS and MFNCSWCC [11].

4.1 The Analysis of the Algorithm Performance

Suppose the spatial data set is S, spatial class set $C=\{C_1, C_2, \dots, C_m\}$ and the instance set $I=\{i_1, i_2, \dots, i_n\}$, the number of the class set C_k is nk ($n=\sum nk$), constraint neighboring class set is t-C-NCS, the length is t ($t < m-1$).

The analysis of the time complexity

The time complexity of the FAMCFNCS includes: computing the correct instance, constructing the NCS database, two-way generating the frequent neighboring class set etc. The time complexity is denoted as $C_n^2 + 2^{m-t} + [(2m-t)(t+1)/2 + (t+1)(t+2)/2]/2$.

The analysis of the spatial complexity

The occupation of the spatial storage variable is closely related to the algorithm and the data structure; the spatial complexity of the FAMCFNCS algorithm can be denoted as $O(\lambda \cdot 2^{m-t})$, λ is a parameter which is related to the length and the support of the constraint neighboring class set.

4.2 Experimental Comparison

We hereby take the algorithms of FAMCFNCS and MFNCSWCC to conduct an analog experiment; suppose there are 12267 correct examples in the spatial data set, its 8178 neighboring class set, the integer which is corresponding to the 8178 neighboring class set is ranging from 3 to 8181(not including 1-NCS), the number of the correct examples neighboring class set which is corresponding to these integer appear as “1”and “2” in turn, we describe it specifically as follows: the correct instance which the corresponding integer 8191 is one, the correct instances which the corresponding integer 8190 are two, there are 13 spatial class sets, represent as {A,B,C,D,E,F,G,H,I,J,K,L,M}, (the sequence number is from right to left), the integer which the constraint frequent neighboring class set corresponding is five, its corresponding frequent neighboring class set is {K, M}, that is the mining constraint premise in the analog experiment.

The experimental circumstances are: the memory is Intel(R) Celeron(R) M CPU 420@1.60 GHz,1G, the operation system is Windows XP Professional, development tool is Visual C# 2005.NET on this platform we can realize the algorithms FAMCFNCS and MFNCSWCC.

As the support of frequent neighboring class set vary, the runtime of the algorithm FAMCFNCS and MFNCSWCC are expressed as Fig.1.

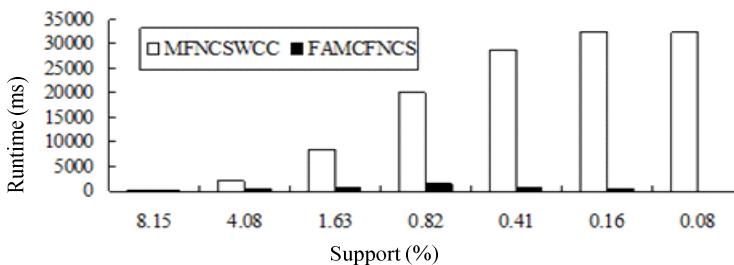


Fig. 1. The Comparison of Runtime as the Support Decrease

As the average length of the frequent neighboring class set varies, the runtime of the FAMCFNCS and MFNCSWCC are expressed as Fig.2.

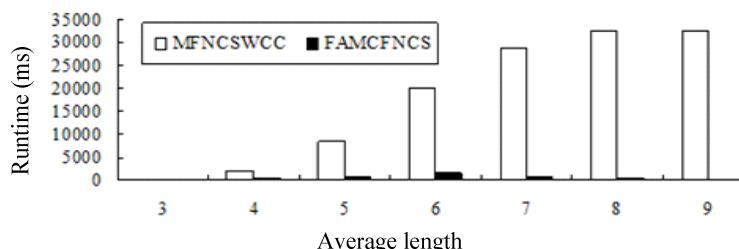


Fig. 2. The Comparison of Runtime as the Average Length Increase

According to these result, we can see that with the decrease of the support, the runtime time of the algorithm FAMCFNCS increase firstly and then decrease, but still faster than the MFNCSWCC. The algorithm FAMCFNCS is suitable for mining any length of the frequent neighboring class set. The effect is much better than algorithm MFNCSWCC.

5 Conclusion

When mining the constraint frequent neighboring class set, because the current algorithm is not suitable for mining the long constraint frequent neighboring algorithm, we put forward a filling algorithm of mining the constraint frequent neighboring class set. It is suitable for the constraint frequent neighboring class set of the any length; the algorithm uses computing subset of neighboring class set to generate candidate via top-down strategy. Meanwhile, it uses the method of filling subset to double generate candidate frequent neighboring class set, and it also uses and operation of corresponding integer to calculate support. The methods which can generate the frequent neighboring class set and compute support quickly will improve our mining efficiency. Through the experiment we prove that it can mine the constraint frequent neighboring class set of the any length.

Acknowledgement. This work is supported by the Key Project Foundation of Sichuan Provincial Education Department (No.11ZA030), Science and Technology Project Foundation of Chongqing Education Commission (No.KJ091108).

References

1. Ma, R., Pu, Y.X., Ma, X.D.: GIS Spatial Association Pattern Ming. Science Press, Beijing (2007)
2. Ma, R., He, Z.-Y.: Mining Complete and Correct Frequent Neighboring Class Sets from Spatial Databases. Geomatics and Information Science of Wuhan University 32(2), 112–114 (2007)
3. Fang, G.: An Algorithm of Alternately Mining Frequent Neighboring Class Set. In: Tan, Y., Shi, Y., Tan, K.C. (eds.) ICSI 2010, Part II. LNCS, vol. 6146, pp. 588–593. Springer, Heidelberg (2010)
4. Zhang, X., Su, F.-Z., Shi, Y.-S., Zhang, D.: Research on Progress of Spatial Association Rule Mining. Progress in Geograph 26(6), 119–128 (2007)
5. Fang, G., Wei, Z., Yin, Q.: An Algorithm of Constrained Spatial Association Rules Based on Binary. In: Sun, F., Zhang, J., Tan, Y., Cao, J., Yu, W. (eds.) ISNN 2008,, Part II. LNCS, vol. 5264, pp. 21–29. Springer, Heidelberg (2008)
6. Fang, G., Wei, Z., Yin, Q.: Extraction of Spatial Association Rules Based on Binary Mining Algorithm in Mobile Computing. In: IEEE Information Conference on Information and Automation, pp. 1571–1575. IEEE Press (2008)
7. Liu, Y.-L., Fang, G.: The research and application of a transaction complement mining algorithm. Journal of Computer Engineering and Applications 44(35), 168–170 (2008)

8. Fang, G., Liu, Y.-L.: Application of Binary System-Based Spatial Mining Algorithm in Mobile Intelligent Systems. *Journal of Southwest University (Natural Science Edition)* 31(1), 95–99 (2009)
9. Fang, G., Xiong, J., Du, X.-L., Tang, X.-B.: Frequent neighboring class set mining. In: 2010 7th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 1442–1445 (2010)
10. Tu, C.-S., Fang, G.: A double search mining algorithm in frequent neighboring class set. In: 2010 IEEE International Conference on Intelligent Computing and Intelligent Systems, pp. 417–420 (2010)
11. Fang, G., Xiong, J., Chen, X.: Frequent Neighboring Class Set Mining with Constraint Condition. In: International Conference on Progress in Informatics and Computing, pp. 242–245 (2010)

Sea Clutter Constant False-Alarm Processing Technology Research Based on Wavelet Transform

Shuang Xu, Yanqiu Cui, and Jifeng Ding

Information & Communication Engineering, Dalian Nationalities University
116600 Dalian, China

Abstract. Constant false-alarm processing is an important digital signal processing method in radar system of the sea clutter automatic detection, which provides the signal detection threshold. Conventional constant false-alarm processing method appears constant false-alarm loss and covers up the detection target in the clutter areas. In order to overcome the shortcomings of traditional constant false-alarm processing, this paper proposes a constant false alarm rate method based on wavelet transform. This method can suppress the interference of sea clutter without impairing the target resolution. Experimental results demonstrate this method can increase the detection accuracy and reduce the false alarm rate.

Keywords: Constant false-alarm, detection threshold, wavelet transform, false alarm rate.

1 Introduction

An automatic identification system (AIS) is a marine traffic monitoring system established by harbor or port authorities, similar to air traffic control for aircraft [1-2]. AIS are designed to improve the safety and efficiency of navigation, safety of life at sea and the protection of the marine environment. The detection and processing of sea clutter is a key part in AIS [3]. Signal detection is to find the target in the noise and clutter interference background. Constant false-alarm processing is an important digital signal processing method in the sea clutter automatic detection. Constant false-alarm processing provides the signal detection threshold, which can avoid the influence of clutter and interference change. The current theory is mainly based on classical test of hypothesis testing statistics in the likelihood ratio test such as ML-CFAR and OS-CFAR[4-5]. Many research institutions have done a lot of research for the selection logic, algorithms and parameters of adaptive determine. The CA-CFAR detector, E-CFAR detector and OSTA-CFAR detector are adaptive detector which are used in different environments [6-7]. In this paper, CA-CFAR detector shown there is a large constant false alarm problem, we propose a wavelet-based CFAR processor.

In order to overcome the shortcomings of traditional constant false-alarm processing, this paper proposes a constant false alarm rate method based on wavelet transform. This method effectively improves the accuracy of radar signal detection, reduces false alarm rate and further improves accuracy of the system.

2 Statistical Models of Clutter Signal

Many distributions have been proposed in the literature to model the amplitude probability density function (PDF) of radar clutter. A large number of experiments show that radar signal mainly submits Weibull distribution and Rayleigh distribution [8].

Rayleigh distribution

$$\text{PDF: } p(u) = \frac{u}{\sigma^2} \exp\left[-u^2/2\sigma^2\right], \quad u \geq 0 \quad (1)$$

Weibull distribution

$$\text{PDF: } p(u) = (\eta/\gamma)(u/\gamma)^{\eta-1} \exp\left[-(u/\gamma)^\eta\right] \quad (2)$$

where η is the shape parameter and γ is the scale parameter.

The Rayleigh PDF is a particular case of the Weibull PDF for $\eta = 2$ and $\gamma^2 = 2\sigma^2$. The low-resolution sea clutter, rain clutter and receiver internal noise in VTS and AIS can be described by Rayleigh distribution.

3 Mean Level CFAR Processing Based on Statistical Properties

In the uniform noise environment, the CFAR detector of ML class can be maintained almost constant according to the selected threshold which limits constant false alarm rate Pf and constant false alarm loss Pd [9-10]. The CFAR detector of ML class usually solves N- square law of averaging around the detect unit. The average is multiplied by a nominal factor T which has been the detection threshold of the detection unit. In the uniform clutter, Pf can detect stable. But in multi-objective and the edge of the interference environment, the detection threshold of great changes will increase rapidly Pf. Figure 1 gives CA-CFAR detector block diagram. Formula (3) is the CA-CFAR detector Pf representation.

$$P_f = \left[1 + \frac{T}{1+\lambda}\right]^{-2n} \quad (3)$$

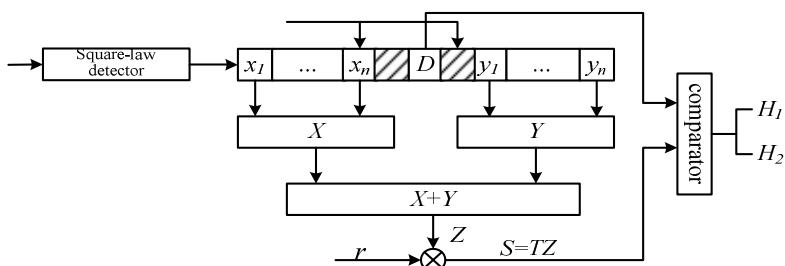


Fig. 1. CA-CFAR processor structure

4 Clutter CFAR Processing Based on Wavelet Transform

In radar signal processing system, false alarm rate is one of the most important technical indicators during signal processing. In order to effectively suppress noise interference and not overload data processor, the radar signal processor is carried out to deal with CFAR. In the real radar signal detection system, the fluctuations of clutter mean average deviating from the statistical average will result in actual decision threshold deviating from the theoretical value. Hence, the actual decision threshold needs to be sure in order to work in a state of constant false alarm.

If it subtracts the smoothing mean value of wavelet transform from radar signal, it can obtain CFAR clutter. The different scale threshold will affect the small target detection of clutter edges and amplitude equalization between great goal and small target. The variance of Radar signal is describes by the details of wavelet transform, and thus this paper gives the soft-threshold algorithm for clutter suppression processing which the threshold is adjusted by the different scale wavelet. It can retain wavelet coefficient of more than absolute value and abandon wavelet coefficient of less than absolute value. Using wavelet coefficient of greater than the threshold restricts the signal. This can remove most of the noise and clutter, at the same time, signal distortion can not obviously caused.

Decomposed with wavelet based on orthogonal compactly supported wavelets, the threshold of the wavelet high frequency coefficient is expressed as

$$\hat{d}^j(n) = \begin{cases} d^j(n), & |d^j(n)| - th_j > 0 \\ 0, & |d^j(n)| - th_j \leq 0 \end{cases} \quad (4)$$

Where the j scale high-frequency wavelet is transform coefficient and th_j is the j scale floating threshold which achieves according to Neymon-Pearson Criterion. It is computed as

$$th_j = \gamma \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d_i^j)^2} \quad (5)$$

where N is the number of points of high frequency wavelet coefficient. The relationship between γ and P_f is shown in table 1.

Table 1. Relationship between γ and P_f

P_f	10^{-7}	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}
γ	5.0	5.0	4.4	3.9	3.3	2.58	1.65

The threshold should multiply the threshold scale control coefficient k in practical applications. Then formula (5) is revised by

$$th_j = k\gamma \sqrt{\frac{1}{N-1} \sum_{i=1}^N (d_i^j)^2} \quad (6)$$

k value is a conversation with the sampling frequency of the radar signal. The value $k=1$ denotes that the sampling frequency is greater or equal than the radar intermediate frequency. Because adjacent sampling point difference of clutter and target becomes larger, the high frequency of radar signal sampling is increased. Wavelet high frequency coefficient is reduced with the increased scale when the sampling frequency is less than the radar intermediate frequency. That is singular characteristic of radar signal. k value is given by

$$k = \frac{2^j}{2^j + 2^{j-1} - 1}, \quad 0 < k \leq 1 \quad (7)$$

where j is the number of scales. Thus, calculating k values are shown in table 2. It can be seen from Table 2 that different scales have different threshold.

Table 2. Threshold control coefficient k

J	1	2	3	4	5	6	7
K	1.0000	0.8000	0.5455	0.3333	0.2130	0.1263	0.0733

The CFAR threshold for the Rayleigh distribution is determined by its statistical average. This paper designs a wavelet transform CFAR (WT-CFAR) processor submitting Rayleigh distribution in the light of the principle. The wavelet transform characteristics of clutter signal demonstrate the feature parameter relationship wavelet transform. The signal $f(t)$ by wavelet transform is given by

$$f(t) = Af_j(t) + \sum_{j=1}^M Df_j(t) \quad (8)$$

where $Af_j(t)$ is the low frequency scale decomposition for $f(t)$ and $Df_j(t)$ is the high-frequency scale decomposition for $f(t)$.

Finding S is the detection threshold.

False alarm probability P_f

$$P_f = \int_S^\infty p(u)du = \int_S^\infty \frac{u}{\sigma^2} \exp[-\frac{u^2}{2\sigma^2}]du = \exp[-\frac{S^2}{2\sigma^2}] \quad (9)$$

Threshold S

$$S = \frac{2E(x)}{\sqrt{\pi}} \sqrt{\ln(1/P_f)} \quad (10)$$

To make $k = \frac{2\sqrt{\ln(1/P_f)}}{\sqrt{\pi}}$, k is control coefficient. When it is $E(x) = \sigma\sqrt{\pi/2}$ and P_f is a given value, CFAR is determined by the threshold achieved formula (6). According to this principle WT-CFAR processor is constructed. WT-CFAR processor structure is shown in Fig. 2.

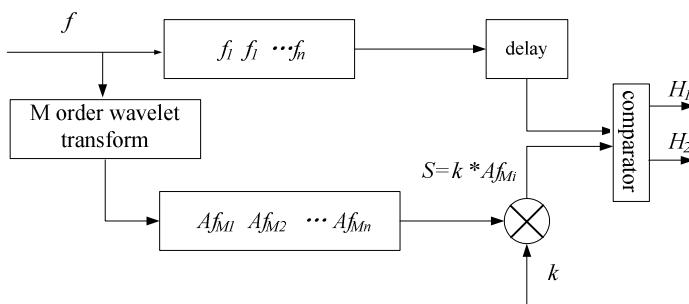


Fig. 2. WT-CFAR processor structure

Where Af_{Mi} is the smooth wavelet coefficient. Comparator output is H_1 when $f \geq S$ and comparator output is H_2 when $f \leq S$.

5 Results and Discussions

5.1 Experiment Result Analysis

WT-CFAR processor can effectively inhibit the clutter of the Rayleigh distribution. The experiment results in DSP are shown in Figure 3. Experiments show that the different false alarm rate has the different clutter suppression. The smaller false alarm rate is the better for clutter suppression, but there will be loss of signal for small target detection. In addition, the clutter CFAR wavelet processor has a very good inhibitory effect for different strength of the clutter.

Figure 3 and Figure 4 show the clutter processing results of the CA-CFAR processor and WT-CFAR processor. It can be concluded that WT-CFAR processor for the Rayleigh distribution of the clutter has a good inhibitory effect.

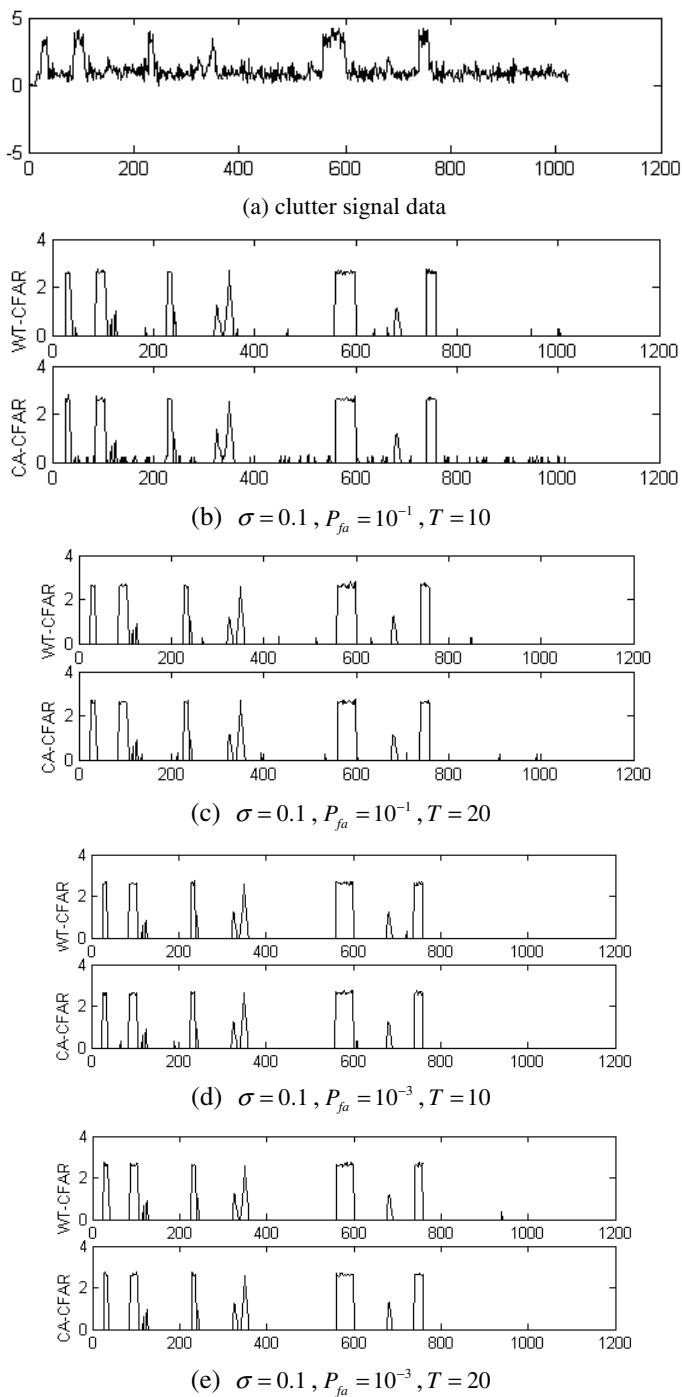


Fig. 3. Comparison of WT-CFAR AND CA-CFAR in variance of 0.1

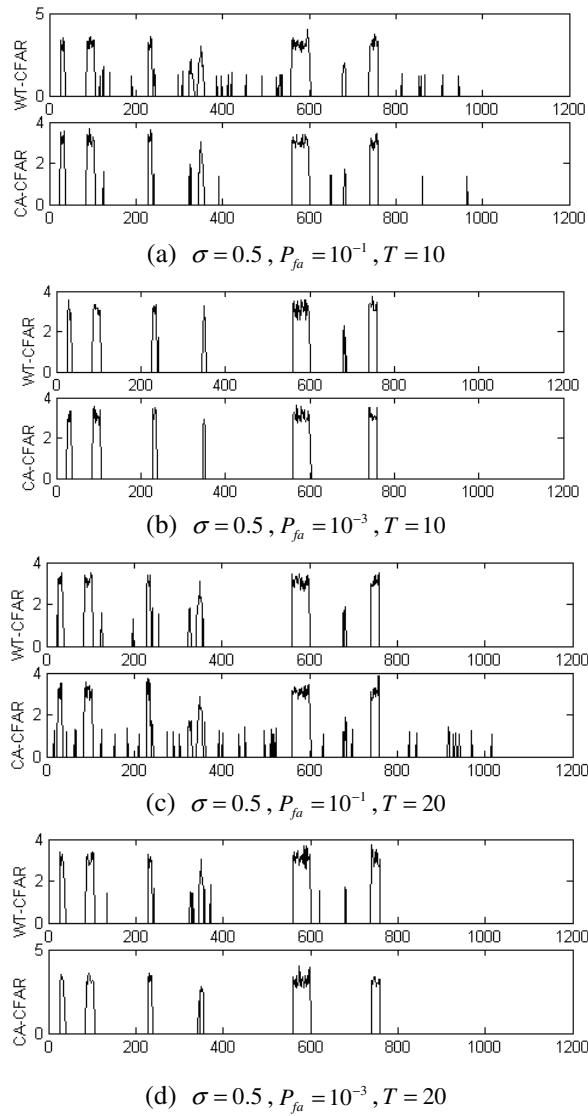


Fig. 4. Comparison of WT-CFAR AND CA-CFAR in variance of 0.1

5.2 Performance Analysis

Evaluation of radar signal is described by power signal noise ratio (SNR).

SNR without clutter

$$SNR = 10 \ln \left(\sum_{i=1}^N [s_0(i)]^2 \right) \Bigg/ \left(\sum_{i=1}^N [n_0(i)]^2 \right) \quad (11)$$

where $s_0(i)$ is the sampled value of the original signal and $n_0(i)$ is the sampled value of the clutter.

SNR with clutter

$$SNR = 10 \ln \left(\sum_{i=1}^N [s_0(i) - n_s(i)]^2 \Big/ \sum_{i=1}^N [n_0(i)]^2 \right) \quad (12)$$

where $n_s(i)$ is the clutter signal and $n_0(i)$ is the sampled value.

According to the formula (12) describing the relationship SNR is shown in the following table 3 under each CFAR checking the 1000 samples.

Table 3. SNR under different false alarm rate

P _f	10-7	10-6	10-5	10-4	10-3	10-2	10-1
SNR/dB	18.20	18.21	18.19	18.18	17.83	17.18	15.62

It can be seen in table 3 that the smaller the false alarm rate is the higher SNR, and vice versa. The false rate is smaller so that the degree of clutter suppression is greater. This is the cause of the higher SNR. Experiments show that different false alarm rate has different SNR and different detection accuracy.

6 Conclusions

According to the shortcomings of traditional constant false-alarm processor in automatic identification system, a constant false alarm rate algorithm based on wavelet transform is put forward. It has been shown by experimentation that the design improves the detection accuracy, reduces the false alarm rate and further improves measurement accuracy of the system.

Acknowledgments. This work was supported by Fundamental Research Funds for the Central Universities (DC110309 and DC110324).

References

1. Pan, J.C., Shao, Z.P., Jiang, Q.S.: Study on Characteristics of Marine Traffic by Data Mining Theory. In: Proceedings of Asia Navigation Conference, pp. 913–917 (2009)
2. Du, C.H.: Research on Mobile AIS Mathematical Model of Offshore Ships Management. Ocean University of China (2008)
3. Zong, P.: Study for Event Mechanism based on AIS. In: Proceedings of 2010 International Colloquium on Computing, Communication, Control, and Management, vol. 4, pp. 301–306 (2010)
4. Wang, X.H., Jin, L., Gao, J.J.: A study on detection algorithm for range-distributed targets in stepped-frequency radar. Modern Radar 5, 35–38 (2009)

5. Sun, Q.Q., Zhao, Y.: Study of Threshold Judgment Coefficient Optimization Based on Constant False Alarm Rate. *Modern Electronics Technique*, 49–52 (2008)
6. Brooker, P.: Airborne collision avoidance systems and air traffic management safety. *Journal of Navigation* 58, 1–16 (2005)
7. Li, K.S.: A study of the effect of melatonin on the proliferation and differentiation of osteoblasts in adolescent idiopathic scoliosis. In: Program and Abstracts of International Conference on Osteoporosis and Bone Research, pp. 525–530 (2008)
8. Chen, Y., Wang, Y.L.D., Zhou, F., Liu, H.: Faint signal processing of lidar based on wavelet multi-resolution analysis. *Laser Technology* 29, 278–283 (2005)
9. Zoheir, H., Faouzi, S.: Distributed IVI-CFAR detection in non-homogeneous environments. *Signal Processing* 84, 1231–1237 (2004)
10. Yu, P., Wang, K.H.: Double dwell hybrid acquisitionin a DS-UWB system. *ETRI Journal* 29, 1–7 (2007)

A Tableau-Based Reasoning Algorithm for Distributed Dynamic Description Logics

Zhuxiao Wang¹, Zhitao Guan¹, Wei Li¹, Kehe Wu¹, Jing Guo², and Guanhua Tian³

¹ School of Control and Computer Engineering, State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, 102206 Beijing, China

² National Computer Network Emergency Response Technical Team/Coordination Center of China, 100029 Beijing, China

³ Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China
{wangzx, guan, liwei, wkh}@ncepu.edu.cn,
guojing.research@gmail.com, tgh_dlut@hotmail.com

Abstract. As a family of dynamic description logics, DDL(X) is constructed by embracing actions into the description logic X, where X represents well-studied description logics ranging from the ALC to the SHIQ. The usage of distributed computation techniques in reasoning is an important premise for the adoption of Dynamic Description Logics in a real-world setting. Practical scalability of DDL(X) reasoning inspired recently a proposal of Distributed Dynamic Description Logics (D3L) framework. D3L is a natural generalization of the DDL(X) framework, which is designed to model the distributed dynamically-changing knowledge repositories interconnected by semantic mappings and to accomplish reasoning in distributed, heterogeneous environments. In this paper, we investigate the reasoning mechanisms in D3L and propose a tableau-based reasoning algorithm for D3L, built as a composition of the state of the art tableaux reasoners for DDL(X).

Keywords: Distributed Reasoning, Dynamic Description Logics, Distributed Dynamic Description Logics, Tableau Algorithms, Inference Engine.

1 Introduction

Description Logics (DLs) are a family of formal knowledge representation languages which structure the knowledge about an application domain in terms of concepts (subsets of individuals in the domain) and roles (binary relations over the domain). Description Logics are playing a central role in knowledge representation, acting as the basis of the well known traditions of Frame-based systems, Semantic Networks and KL-ONE-like languages, Object-Oriented representations, Semantic data models, and Type systems [1-5].

By introducing a dynamic dimension into the description logics, Shi et al [6], [7] propose a family of Dynamic Description Logics named DDL(X) for uniformly

representing and reasoning about dynamic application domains [8], [9], where X represents well-studied description logics ranging from the \mathcal{ALC} to the \mathcal{SHOIQ} .

Scalability is an issue that needs to be considered when designing any reasoner for dealing with large and complex ontologies and large data sets. The practical usage of distributed computation techniques in reasoning is an important premise for the adoption of Dynamic Description Logics in a real-world setting. In this paper we describe a novel approach of applying distributed computation techniques to DDL(X) reasoning. The approach is to design a logical framework of Distributed Dynamic Description Logics (D3L) capable of capturing the dynamic behavior of the overall system. D3L is a natural generalization of the DDL(X) framework, which is designed to model the distributed dynamically-changing knowledge repositories interconnected by semantic mappings and to accomplish reasoning in distributed, heterogeneous environments. Afterwards, we present a tableau-based distributed reasoning procedure for providing the capability of global reasoning in D3L and decomposing large reasoning tasks to sub-tasks that could be concurrently processed by different local reasoning engines.

Our contributions in this paper are as follows: (i) we define the semantics and syntax of D3L to formally capture the dynamic behavior of the overall system; (ii) analyze semantical mechanisms allowing for propagating the dynamic knowledge, i.e. how dynamic knowledge propagates through local reasoning engines; and finally, (iii) propose a distributed tableau algorithm for providing the capability of global reasoning in such a system and decomposing large tasks to sub-tasks that could be tackled by different local reasoning engines.

In the following sections, we firstly present the Distributed Dynamic Description Logics (D3L) framework and enunciate the main properties in Section 2. Furthermore, in Section 3 we describe the abstract distributed tableau algorithm that computes action subsumption in D3L. We show how local tableaux are combined in order to implement global reasoning services in D3L. Afterwards, we show how a limited case of D3L is handled by the approach. Finally, we summarize the paper in Section 4.

2 The Distributed Dynamic Description Logics Formalism

In this section, we present the basic definitions of the Distributed Dynamic Description Logics (D3L) formalism. From a theoretical perspective, the D3L is based on the long tradition of logics for distributed systems, and based on extensions to Dynamic Description Logics introduced in [6], [7]. If we do not consider the dynamic dimension of D3L, D3L can be reduced to Distributed Description Logics [10], [11]. Let I be a nonempty set of indexes, and DDL $_i$ be dynamic description logics for every $i \in I$. A sequence $D3L = \{DDL_i\}_{i \in I}$ is then called a distributed dynamic description logic. We label each description C in DDL $_i$ with its index i (written as $i:C$) to indicate that some description C belongs to the language of the dynamic description logic DDL $_i$. Three types of assertions—concept bridge rules, action bridge rules and individual correspondence—are used to express relations between the components of a Distributed Dynamic Description logic.

Definition 1. Given concepts (or actions) C and G of DDL_i and DDL_j respectively, a bridge rule is an expression of the form

$$i:C \xrightarrow{\subseteq} j:G \quad \text{into-bridge rule} \quad (1)$$

or of the form

$$i:C \xrightarrow{\supseteq} j:G \quad \text{onto-bridge rule} \quad (2)$$

If C and G are concepts of DDL_i and DDL_j respectively, into- and onto-bridge rules are called concept bridge rules. If both C and G are actions, bridge rules are called action bridge rules.

Let x be an individual of DDL_i and y individuals of DDL_j. An individual correspondence is an expression of the form

$$i:x \mapsto j:y \text{ individual correspondence} \quad (3)$$

A distributed TBox (DTBox) DT = <{Ti}i ∈ I, P> consists of a collection of T-boxes {Ti}i ∈ I and a set P = {Pij}i ≠ j ∈ I of concept bridge rules. A distributed ABox (DABox) DA = <{Ai}i ∈ I, C> consists of a collection of A-boxes {Ai}i ∈ I together with a set C = {Cij}i ≠ j ∈ I of individual correspondences. A distributed ActBox (DActBox) DAct = <{Acti}i ∈ I, H> consists of a collection of ActBoxes {Acti}i ∈ I and a set H = {Hij}i ≠ j ∈ I of action bridge rules. A distributed dynamic knowledge base is a triple K = (DT, DA, DAct).

The semantics of a Distributed Dynamic Knowledge Base (DDKB) is formally defined as follows.

Definition 2. A distributed model M of a DDKB K = (DT, DA, DAct) is a tuple <{Mi = (Wi, Ti, Δ^{Ii}, II)}i ∈ I, {rij}i ≠ j ∈ I, {stateij}i ≠ j ∈ I>, where each Mi is a local model for the corresponding DDLi on local domains Δ^{Ii} and a domain relation rij from Δ^{Ii} to Δ^{Ij} is defined as a subset of Δ^{Ii} × Δ^{Ij}. Given a point d ∈ Δ^{Ii} and a subset D ⊆ Δ^{Ii}, we set

$$r_{ij}(d) = \{d' \in Δ^{Ij} \mid (d, d') \in r_{ij}\}, \quad r_{ij}(D) = \bigcup_{d \in D} r_{ij}(d) \quad (4)$$

A state relation state_{ij} from W_i to W_j is defined as a subset of W_i × W_j. Given a point w ∈ W_i and a subset T_i(α) ⊆ (W_i × W_i), we set

$$state_{ij}(w) = \{w' \in W_j \mid (w, w') \in state_{ij}\} \quad (5)$$

$$state_{ij}(T_i(\alpha)) = \bigcup_{(w, w') \in T_i(\alpha)} state_{ij}(w) \times state_{ij}(w') \quad (6)$$

With respect to any state w ∈ W_i and z ∈ W_j, a distributed model M is said to d-satisfy (written (M, w, z) ⊨ d) concept bridge rules and individual correspondences according to the following clauses:

$$(M, w, z) \models d \quad i:A \xrightarrow{\subseteq} j:F \text{ iff } r_{ij}(A^{I_i(w)}) \subseteq F^{I_j(z)}, \text{ concept into-bridge rule} \quad (7)$$

$$(M, w, z) \models d i : B \xrightarrow{\cong} j : G \text{ iff } r_{ij}(B^{I_i(w)}) \supseteq G^{I_j(z)}, \text{concept onto-bridge rule} \quad (8)$$

$$(M, w, z) \models d i : x \mapsto j : y \text{ iff } y^{I_j(z)} \in r_{ij}(x^{I_i(w)}) \text{ . individual correspondence} \quad (9)$$

Secondly, the satisfaction of an action bridge rule br in M, written as $M \models d br$, is defined as follows:

$$M \models d i : \alpha \xrightarrow{\subseteq} j : \beta \text{ iff } state_{ij}(T_i(\alpha)) \subseteq T_j(\beta), \text{action into-bridge rule} \quad (10)$$

$$M \models d i : \pi \xrightarrow{\cong} j : \rho \text{ iff } state_{ij}(T_i(\pi)) \supseteq T_j(\rho). \text{action onto-bridge rule} \quad (11)$$

The $M \models d$ is standard for formulas of the component Dynamic Description Logics [7].

As usual, $DT \models d i : C \sqsubseteq D$ means that for every distributed model M, $M \models d DT$ implies $M \models d i : C \sqsubseteq D$. The same definition could be applied to ABoxes DA, ActBoxes DAct and individual assertions.

Let us see now how action bridge rules affect the forward propagation of dynamic knowledge in D3L. The basic idea preceding that result is that combination of action onto- and into-bridge rules allows for directional propagating the action knowledge across knowledge repositories in form of DDL(X) action subsumption axioms [6], [7].

A changing world can be viewed as an entity that changes over time. At each point in time, the world is in one particular state that determines how the world is perceived by the entities acting in the world. By means of well-defined change operations named actions, any action can affect the world and modify the world state if its preconditions are satisfied. Such actions denote state transitions in the set of states.

Theorem 1 (Sequence action propagation). If H_{ij} contains $i : \alpha \xrightarrow{\subseteq} j : \beta$ and $i : \pi \xrightarrow{\subseteq} j : \rho$, then:

$$DAct \models d i : \alpha ; \pi \Rightarrow DAct \models d j : \beta ; \rho \quad (12)$$

where α , π , β , and ρ are actions.

Theorem 2 (Simple action subsumption propagation). Combination of onto- and into-bridge rules allows to propagate subsumptions across knowledge repositories. Formally, if H_{ij} contains $i : \alpha \xrightarrow{\subseteq} j : \beta$ and $i : \pi \xrightarrow{\cong} j : \rho$, then:

$$DAct \models d i : \pi \sqsubseteq \alpha \Rightarrow DAct \models d j : \rho \sqsubseteq \beta \quad (13)$$

Theorem 3 (Generalized action subsumption propagation). If H_{ij} contains $i : \pi \xrightarrow{\cong} j : \rho$ and $i : \alpha k \xrightarrow{\subseteq} j : \beta k$ for $i \leq k \leq n$ ($n \geq 0$), then:

$$DAct \models d i : \pi \sqsubseteq \bigcup_{k=1}^n \alpha k \Rightarrow DAct \models d j : \rho \sqsubseteq \bigcup_{k=1}^n \beta k \quad (14)$$

The theorem 3 is important as it constitutes the main reasoning step of the tableau algorithm proposed in the next section. Given the limited space available, in this article I will not delve into the details of the proofs of the above properties.

3 Distributed Reasoning in D3L

Algorithm 1 implements a distributed reasoning procedure (Table 1). Here we define a distributed tableau procedure DDTableaux_j (one for each $j \in I$), which takes as an input a formula φ to be verified and returns the result of its (un) satisfiability test. DDTableaux_j tries to build a model of $M_j = (W_j, T_j, \Delta^{I_j}, I_j)$ with $(M_j, w) \models \varphi$, using the expansion rules in [7], plus an additional “ DiH_{ij} -rule” expansion rule (Table 2).

Example 1. For a distributed ActBox $\mathcal{D}\mathcal{A}ct_{ij} = \langle \text{Act}_i, \text{Act}_j, \mathcal{H}_{ij} \rangle$, suppose that Act_i contains axioms $\pi_1 \sqsubseteq \alpha_1$ and $\pi_2 \sqsubseteq \alpha_2$, Act_j does not contain any axiom, and that \mathcal{H}_{ij} contains the following action bridge rules:

$$i: \alpha_1 \xrightarrow{\sqsubseteq} j: \beta_1 \quad i: \alpha_2 \xrightarrow{\sqsubseteq} j: \beta_2 \quad (15)$$

$$i: \pi_1 \xrightarrow{\exists} j: \rho_1 \quad i: \pi_2 \xrightarrow{\exists} j: \rho_2 \quad (16)$$

Table 1. The distributed satisfiability-checking algorithm

Algorithm 1 Distributed reasoning procedure

$\text{DDTableaux}_j(\varphi)$

- 1: Construct a branch $\mathcal{B} := \{ \sigma_0, \varepsilon_0; \text{nf}_{\mathcal{T}, \mathcal{A}}(\varphi) \}$
 - 2: If tableau expansion rules in [7] and Table 2 can be applied to \mathcal{B} in such a way that they yield a completed branch \mathcal{B}' , and
 - \mathcal{B}' is neither contradictory nor ignorable, and
 - the initial view $IV_{\mathcal{B}'}$ is consistent w.r.t. \mathcal{R} and \mathcal{T} ,
then the algorithm returns “ φ is satisfiable”, and “ φ is unsatisfiable” otherwise.
-

Table 2. The action bridge expansion rule

Rules	Definitions
DiH_{ij} -rule	If (1) $\sigma, \varepsilon: T_j(\rho)((w, w')) \in B$, such that $i: \pi \xrightarrow{\exists} j: \rho \in H_{ij}$, and $i: \alpha_k \xrightarrow{\exists} j: \beta_k \in H_{ij}$ for $1 \leq k \leq n$, $n \geq 0$, and (2) the procedure $\text{DDTableaux}_i(\phi')$ decides that the formula $\phi' = (Ti(\pi) \cap \cup_{1 \leq k \leq n} Ti(\alpha_k)) \cap ((z, z'))$ is unsatisfiable w.r.t. the i -th dynamic description logics knowledge base, where we assume that an state transition (z, z') is an instance of an action $Ti(\pi) \cap \cup_{1 \leq k \leq n} Ti(\alpha_k)$ in the i -th dynamic description logics, for some $\sigma, \varepsilon: \cup_{1 \leq k \leq n} Tj(\beta_k) \notin B$, then set $B := B \cup \{\sigma, \varepsilon: \cup_{1 \leq k \leq n} Tj(\beta_k) ((w, w'))\}$.

Let us show that $DAct_{ij} \vDash j: \rho_1 \cup \rho_2 \sqsubseteq \beta_1 \cup \beta_2$, i.e. that for any distributed interpretation $M = < M_i = (W_i, T_i, \Delta^{I_i}, I_i), M_j = (W_j, T_j, \Delta^{I_j}, I_j), r_{ij}, state_{ij}, T_j(\rho_1 \cup \rho_2) \sqsubseteq T_j(\beta_1 \cup \beta_2)$.

1. Suppose that by contradiction there is an $(w, w') \in W_j \times W_j$ such that $(w, w') \in T_j(\rho_1 \cup \rho_2)$ and $(w, w') \notin T_j(\beta_1 \cup \beta_2)$.
2. Then either $(w, w') \in T_j(\rho_1)$ or $(w, w') \in T_j(\rho_2)$, and $(w, w') \notin T_j(\beta_1), (w, w') \notin T_j(\beta_2)$.
3. Let us consider the case where $(w, w') \in T_j(\rho_1)$. From the fact that $(w, w') \notin T_j(\beta_1)$, by the bridge rule (15), there is $(z, z') \in W_i \times W_i$ with $\langle(z, z'), (w, w')\rangle \in state_{ij}$, such that $(z, z') \notin T_i(\alpha_1)$.
4. From the fact that $(w, w') \in T_j(\rho_1)$, by bridge rule (16), we can infer that for all $(z, z') \in W_i \times W_i$ if $\langle(z, z'), (w, w')\rangle \in state_{ij}$ then $(z, z') \in T_i(\pi_1)$.
5. But, since $\pi_1 \sqsubseteq \alpha_1 \in Act_i$, then $(z, z') \in T_i(\alpha_1)$, and this is a contradiction.
6. The case where $(w, w') \in T_j(\rho_2)$ is similar.

The above combination of a tableau in Act_j with a tableau in Act_i gives a distributed tableau in \mathcal{DAct}_{ij} , depicted in Fig. 1.

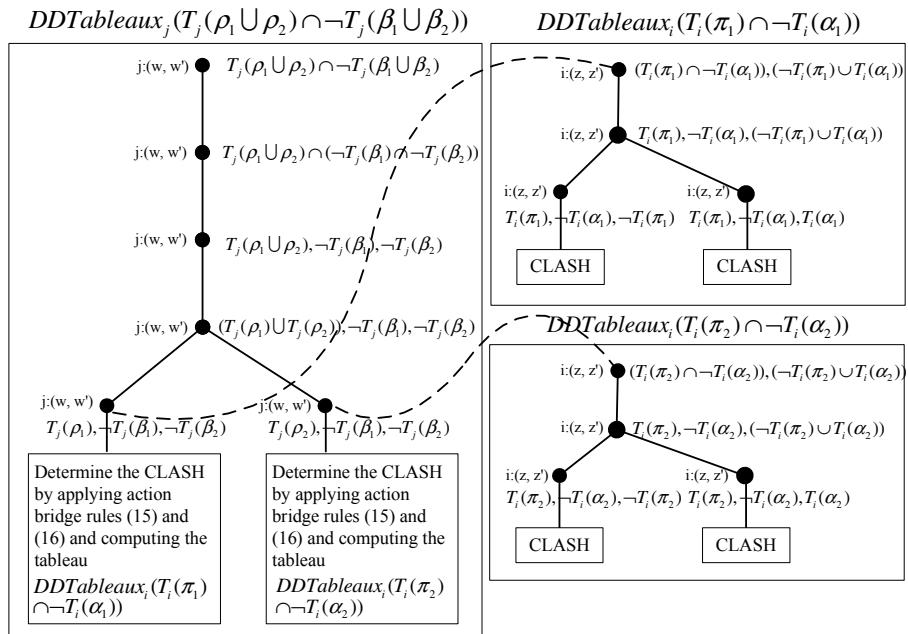


Fig. 1. Graphical intuition of distributed Tableau-based D3L reasoning.

4 Conclusions

The last decade of basic research in the area of Dynamic Description Logics (DDL(X)) has created a stable theory, efficient inference procedures, and has demonstrated a wide applicability of DDL(X) to dynamic knowledge representation and reasoning. Usage of distributed computation techniques in reasoning for DDL(X) establishes a major research problem. The approach proposed in this paper is to design a logical framework of Distributed Dynamic Description Logics (D3L), which is composed of a set of stand alone DDL(X) pairwise interrelated with each other via collection of bridge rules. Furthermore, we present a tableau-based distributed reasoning procedure for providing the capability of global reasoning in D3L and decomposing large reasoning tasks to sub-tasks that could be concurrently processed by different local reasoning engines.

The paper represents a work in progress. Thus it has many open issues for the future research directions. In recent years distributed computing architectures emerge, such as grid [12], multi-agent [13], peer-to-peer [14], or multi-core machines [15]. This allows for the overall computational power to grow further, provided that tableau-based reasoning algorithms respect this computational paradigm shift. Furthermore, it is also a valuable and interesting work to combine the proposed approach seamlessly with these distributed computing architectures.

Acknowledgments. This work is supported by the Fundamental Research Funds for the Central Universities (No.11QG13) and the National Science Foundation of China (No.71101048).

References

1. Bonatti, P., Lutz, C., Wolter, F.: Description logics with circumscription. In: Proc of the 10th Int. Conf. on Principles of Knowledge Representation and Reasoning, pp. 400–410. AAAI Press, Menlo Park (2006)
2. Jiang, Y., Wang, J., Tang, Y., Deng, P.: Semantics and reasoning of description logic μ ALCQO. Journal of Software 20, 491–504 (2009)
3. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P.: The Description Logic Handbook: Theory, Implementation and Applications. Cambridge University Press, Cambridge (2002)
4. Horrocks, I.: DAML+OIL: A description logic for the semantic web. Bull of the IEEE Computer Society Technical Committee on Data Engineering 25, 4–9 (2002)
5. Horrocks, I., Patel-Schneider, P.F., Harmelen, F.V.: From SHIQ and RDF to OWL: the making of a web ontology language. Journal of Web Semantics 1, 7–26 (2003)
6. Shi, Z., Dong, M., Jiang, Y., Zhang, H.: A logical foundation for the semantic web. Science in China, Ser. F 48, 161–178 (2005)
7. Chang, L., Shi, Z., Gu, T., Zhao, L.: A Family of Dynamic Description Logics for Representing and Reasoning About Action. J. Autom. Reasoning, 1–52 (2010)
8. Wang, Z., Yang, K., Shi, Z.: Failure Diagnosis of Internetwork Systems Using Dynamic Description Logic. J. Softw. China 21, 248–260 (2010)

9. Chen, L., Hu, H., Shi, Z.: Reasoning about Web Services with Local Closed World Assumption. In: Proc. of Int. Conf. Web Intelligence, pp. 367–370. IEEE Press, New York (2009)
10. Borgida, A., Serafini, L.: Distributed Description Logics: Assimilating Information from Peer Sources. In: Spaccapietra, S., March, S., Aberer, K. (eds.) *Journal on Data Semantics I*. LNCS, vol. 2800, pp. 153–184. Springer, Heidelberg (2003)
11. Kutz, O., Lutz, C., Wolter, F.: Epsilon-connections of abstract description systems. *Artificial Intelligence* 156, 1–73 (2004)
12. Krauter, K., Buyya, R., Maheswaran, M.: A taxonomy and survey of grid resource management systems for distributed computing. *Software-Practice & Experience* 32, 135–164 (2002)
13. Shi, Z., Zhang, H., Dong, M., Zhao, Z.: MAGE: multi-agent environment. In: 2nd Int. Conf. on Computer Networks and Mobile Computing, pp. 181–188. IEEE Computer Soc., Shanghai (2003)
14. Androultsellis-Theotokis, S., Spinellis, D.: A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys* 36, 335–371 (2004)
15. Pham, D.C., Aipperspach, T., Boerstler, D., Bolliger, M., Chaudhry, R.: Overview of the architecture, circuit design, and physical implementation of a first-generation cell processor. *IEEE Journal of Solid-State Circuits* 41, 179–196 (2006)

Improvement of Target Extraction and Dense Matching

Zetao Jiang, Qiang Wang, and Yanru Cui

Information Engineering School, Nanchang Hangkong University, Nanchang, China
{zetaojiang,wangqiang000ws}@126.com, cuiyanrumwyauavk@163.com

Abstract. In order to obtain dense 3D point clouds of the target based on image sequence, it is necessary to ensure the accuracy of the target extraction and the dense of the disparity which based on the dense matching of image sequence target. A new method that was used to improve the target extraction and dense matching was proposed in this paper. In the image target extraction, many interactive iteration segmentations method was adopted to amend the inaccurate region that was segmented by traditional GrabCut; and in the dense matching, the image texture features information was introduced to divide the image into the texture-rich region and sparse texture region, and different dense matching methods were applied to the different regions. This method solved the problem that image target can not be accurately extracted from a complex scene; On the other hand, the localized reconstruction due to the fact that image texture sparse region can not be achieved quasi-dense matching was also solved. The results displayed that, after the using of this method, good results have been showed, and the accuracy of three-dimensional reconstruction was improved.

Keywords: image segmentation, texture analysis, dense matching.

1 Introduction

The image object extraction is the use of image segmentation method to extract useful target in the image. The traditional image segmentation methods can be divided into three categories: segmentation method based on threshold, segmentation method based on edge and segmentation method based on region. Nowadays, interactive technologies take advantage of the awareness of people for the object to find the boundary and obtain high performance segmentation through appropriate interactive. Classical methods include Snakes [1], Intelligent Scissors [2], GraphCut [3] and so on. The article is the use of many interactive iteration segmentations method to amend the inaccurate region that has been segmented by traditional GrabCut. and ultimately the accurate target extraction of complex scenes images.

In the field of computer vision, the results of image matching will direct impact the subsequent 3D reconstruction. In order to achieve the 3D reconstruction of image sequences, it is necessary to ensure the accuracy and efficiency of the Image dense matching. The earlier proposed dense stereo matching method is a window matching algorithm [4] which relied on the window constraints, the main idea of their method

use the gray of adjacent pixel in a window to compare, and use its similarity to determine whether match with the center of the window. The quality of the method results depends on the size of the window. Selecting the optimal window size which is determined by the image texture and local variation of parallax, what is more the best window in the experiment is often difficult to determine. In modern times, many domestic and foreign scholars have been a lot of research on the dense stereo matching algorithm. Collaborative iterative algorithm [5] using global constraints and iterative algorithm to get a better disparity map. But the algorithm uses a variety of algorithms, so the Computational complexity and taking up storage space. Dense matching method based on triangulation uses the specific characteristics of a triangle to limit the geometric structure of objects. This algorithm can narrow the scope of the matching, thereby improving the speed and accuracy of the matching.

The above dense matching methods reflect the good performance, but some algorithms only have good effect in the texture-rich region, the dense matching is not effective in the texture sparse region. To solve this problem, this paper presented an improved dense stereo matching method: different dense matching methods were applied to the different regions. The method can solve the problem of the stereo matching in sparse texture region and improve the accuracy of the disparity map.

2 Improved Image Object Extraction Based on GrabCut

Traditional GrabCut algorithm is sensitive to the changes of image brightness, so it will be inaccurate target objects for GrabCur to target extraction in span brightness variations of the image [6].

In order to solve the lack of traditional GrabCut algorithm, this paper presented an improved method based on GrabCut, which was the use of multiple mouse interactions and iteration segmentations method to amend the inaccurate region that had been segmented by traditional GrabCut. The improved method is as follows:

1) First of all, the initial target was segmented by GrabCut and calculated the contour area of the initial target; Followed by, the mouse again interacted region which had been interacted by the initial mouse, the interaction region together with the pre-target object contour area calculated the region which did not meet user requirements. Last, the interaction region and the previous contour line together formed the correction area, used to amend the inaccurate region based on GrabCut;

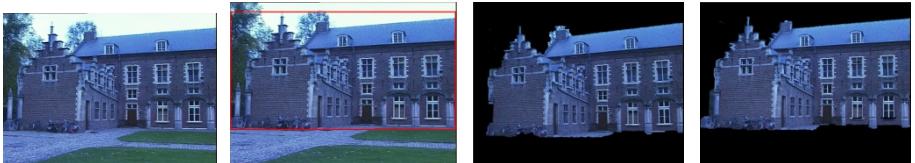
2) According to the above of the foreground and background model remapped the correction area to s-t network, then amended image by minimum cost cutting iterative method;

3) Output corrected results until iterative minimize.

This paper used traditional GrabCut segmentation method and improvement of GrabCut to achieve object extraction. The experimental results are as follows:



(a) The original image (b) Target selection (c) Before target corrected (d) After target corrected

Fig. 1. The first experiment: the target correction

(a) The original image (b) Target selection (c) Before target corrected (d) After target corrected

Fig. 2. The second experiment: the target correction

(a) The original image (b) Target selection (c) Before target corrected (d) After target corrected

Fig. 3. The third experiment: the target correction

Figure 1 (c), 2 (c), 3 (c), 4 (c) is the result that is before target corrected, Figure 1 (d), 2 (d), 3 (d), 4 (d) is the result that is target corrected. Compared the four groups of target extraction, the result confirm that the correction method is simple. This method is adopted to amend the inaccurate region that is segmented by traditional GrabCut.

3 The Improved Dense Matching Algorithm Based on Regional Growth

Following the target extraction based on image sequence with improved GrabCut, sparse matching was performed by feature matching based on SURF [7], and the situation that a lot of match points appeared in the sparse region occurred, little match points appeared in the texture-rich region. The image sequence is made up with texture-rich region and the texture sparse region, and existed sparse matching methods are difficult to solve regional matching of texture sparse, so the dense matching for the target of image sequence using traditional dense matching based on regional growth will make the texture sparse region matching relations stop, and ultimately affect the accuracy of reconstruction. To solve this problem, an improved

dense matching method was adopted. The method uses a two-step strategy: Dense matching in the texture-rich region and sparse region [8]. Many fine matching points were obtained by feature matching based on SURF in the texture-rich region, so dense matching would be carried out using dense matching based on regional growth [9]. After the calculation of the feature information of points in texture sparse region, the points which have obvious characteristics were chosen as key points, then the key points were seed points, Quasi-dense matching was conducted based on regional growth.

3.1 Sparse Texture and Texture-Rich Area Segmentation Based on Texture Features from Image

GLCM (gray levelco-occur-rence matrix, GLCM) is a commonly used method to describe the texture through the study of gray space-related features, which has high value in the research and analysis of texture images. Therefore, the GLCM was used to describe the image texture features in this paper.

GLCM is expressed by the joint probability density through two different pixels, which not only reflects the light distribution characteristics, but also reflects the characteristics of the location distribution among pixels of the same intensity or similar intensity. GLCM is the basis to constitute texture features, which has the Second-order statistical properties of the image brightness changes. Since GLCM has the characteristics of computational complexity, in order to reduce the complexity of the algorithm of extracting image texture features. It can take use of the Symmetry characteristics of GLCM, and select four directions ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) for operating, This makes the calculation of the entire algorithm greatly reduced. The GLCM calculation formula used in this article is:

$$P(i, j) = \sum_{\theta=0^\circ, 45^\circ, 90^\circ, 135^\circ, d=1} P(i, j, d, \theta) \quad (1)$$

Access to a lot of literatures, the following four statistics were most commonly used to extract image texture features, which has the best results:

(1) Angular second moment

$$f_{ASM} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} [P(i, j)]^2 \quad (2)$$

Angular second moment is the square of the various elements of GLCM, which is also known as energy. It is the uniform measure of the image texture gray, reflecting the degree of the uniformity of gray level distribution and texture coarseness.

(2) Contrast

$$f_{CON} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 P(i, j) \quad (3)$$

Contrast is the inertia moment of the GLCM near the main diagonal, which measures how the value of the matrix distributed, and the local changes of the image, reflecting the image clarity and grooves depth of texture.

(3) Relevant

$$f_{CORRLN} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} ijP(i, j) - \mu_i \mu_j / \sigma_i^2 \sigma_j^2 \quad (4)$$

It measures the similarity of the elements in the spatial gray level co-occurrence matrix on the row or column direction; therefore, the relevant values reflect the correlation of local gray-scale in the image.

(4) Entropy

$$f_{ENT} = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (-\ln P(i, j)) P(i, j) \quad (5)$$

Entropy measures the randomness of the texture. It is maximum when all values of the space GLCM are equal; conversely, if the values of GLCM are very uneven, its value is small.

It can be seen from the above (2) - (5), when the value of f_{CON} , f_{CORRLN} and f_{ENT} is smaller, gray scale of pixels is identical or similar in this region, namely, that the region is the sparse texture region.

In order to achieve the texture divided, This paper set the threshold for f_{CON} , f_{CORRLN} and f_{ENT} , and ensured that the Image was divided into rich texture and sparse texture regions. Here, the threshold of f_{CON} , f_{CORRLN} and f_{ENT} was set as t_1 , t_2 and t_3 respectively (the values of the introduction threshold t_1 , t_2 and t_3 depend on the texture feature of the specific image). Then, when they satisfied the following conditions: $f_{CON} < t_1$, $f_{CORRLN} < t_2$, $f_{ENT} < t_3$, the image region was a sparse texture region, otherwise, it was a texture-rich region.

Then, This paper used the above method to mark texture-rich region and sparse region. The results were shown in the following figure 6, figure 7 and figure 8.



(a) The targets extraction (b) The texture-rich and sparse texture regions are marked

Fig. 6. The first experiment: the texture-rich and sparse texture regions are marked



(a) The group of the target are extracted (b) the texture-rich and sparse texture regions are marked

Fig. 7. The second experiment: the texture-rich and sparse texture regions are marked



(a) The group of the target object are extracted (b) the texture-rich region and single texture regions are marked

Fig. 8. The third experiment: the texture-rich and sparse texture regions are marked

The red curve in the above experimental results divides the object of the image into many rich-texture regions and sparse-texture regions. The results show that, by using the method in this article, the images are marked as rich-texture region and sparse-texture region.

3.2 Matching of the Texture-Rich Region and Sparse Texture Region

By the above Method, image target was divided into rich texture region and sparse texture region. This paper used the dense matching based on regional growth for the texture-rich region .In the sparse texture region, firstly , through points which was defined as the seed points to build stereo matching relations, secondly , used the seed points as original point to achieve quasi-dense matching based on the regional growth method . By this way, the feature points mathching from few or no to quasi-dense matching in the sparse texture region.

In the sparse region of image texture , this article described feature points based on the synthesis of color features information and texture feature information .the feature points can be described as $V_p = (fCON, fCORRLN, fENT, R, G, B) = (T, C)$,in this formula, $fCON$, $fCORRLN$, $fENT$ are texture eigenvalue , R , G , B ,the three character express three channel of color. $T=(fCON, fCORRLN, fENT)$ is the V_p vector of texture feature, $C=(R, G, B)$ is color vector. In texture sparse regions define the most remarkable feature of the point as the key point. The paper defined the comprehensive texture feature and color characteristics which are the most obvious as can be used as the key point of the sparse texture region. That is,

$$d = \sqrt{|T|^2 + |C|^2} \quad (6)$$

When d is the maximum, that point can be used as the key point [10]. The key points in sparse texture region are shown in map 9(a),10(a),11(a), then used the NCC method to establish the matching relation between key points in image sequence, And then used the key points as seed points to achieve dense matching based on regional growth . This paper achieved the sparse texture feature points region from few or no to quasi-dense matching.

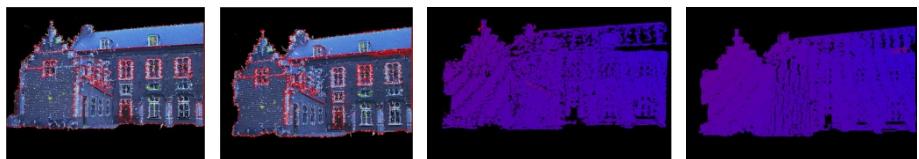
The above method uses a two-step strategy to achieve image matching: Dense matching in the texture-rich region and sparse region. The dense disparity map can be obtained by this method, which provide a good foundation for the subsequent three-dimensional reconstruction. The experimental results are as follows:



(a) The set key points (b) The disparity map of the dense (c) the disparity map matching based on regional growth of this algorithm

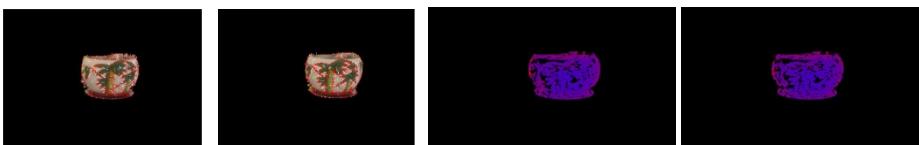
Fig. 9. The disparity maps of this algorithm and dense matching based on regional growth

Clover-like color points are the seed point of the texture-rich region and the key point of sparse texture region are shown in Figure 9(a), 10(a), 11(a). The disparity map of the traditional dense matching based on regional growth are shown in Figure 9(b), 10(b), 11(b), the final disparity map of this algorithm are shown in Figure 9(c), 10(c), 11(c).



(a) The set key points (b) The disparity map of the dense (c) the disparity map matching based on regional growth of this algorithm

Fig. 10. The disparity maps of this algorithm and dense matching based on regional growth



(a) The set key points (b) The disparity map of the dense (c) the disparity map matching based on regional growth of this algorithm

Fig. 11. The disparity maps of this algorithm and dense matching based on regional growth

Comparison of experimental results by the above three groups confirm that this method obtain can obtain the more dense disparity map. The proposed method improved the dense matching based on regional growth which can not achieve image matching in the sparse texture region. So it provides a good basis for the follow-up dense three-dimensional surface information.

4 Conclusion

In order to achieve the accurate target extraction, this paper presented an improved GrabCut algorithm for image object extraction, many interactive iteration segmentations method was adopted to amend the inaccurate region that was segmented by traditional

GrabCut. Experimental results show that this method can amend region that was segmented by GrabCut, so that the revised target was more accurate.

In order to solve the problem of stereo matching of sparse texture region, this paper presented an improved dense matching approach, which uses a two-step strategy: in the texture-rich region, use the existing dense matching based on regional growth for dense matching of the texture-rich region; in sparse texture region, and use regional growth method based on the key point for dense matching. The localized reconstruction due to the fact that image texture sparse region can not be achieved quasi-dense matching was also solved. Finally, this paper did multiple sets of experiments, the results of experiments confirm the feasibility and accuracy of this method, which has practical value in the field of computer vision.

Acknowledgments. This paper is sponsored by Nature Science Foundation of China (60973096) and Science Foundation of aviation in China (2010ZC56007).

References

1. Sung, W.Y., Hai, K.L., Jeong, H.K.: Medical Endoscopic Image Segmentation Using Snakes. IEICE Transactions on Information and Systems E87-D(3) (2004)
2. Suetake, N., Uchino, E., Hirata, K.: Separability-Based Intelligent Scissors for Interactive Image Segmentation. IEICE Transactions on Information and Systems E90-D(1) (2007)
3. Rother, C., Kolmogorov, V., Blake, A.: “GrabCut”-Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics 23(3) (2004)
4. Yuan, M., Liu, Z., Wang, F.: Improved algorithm for local weighted matching cost computation. Journal of Image and Graphics 16(9), 1664–1669 (2011)
5. Zheng, Z.-G., Wang, Z.-F.: A Region Based Stereo Matching Algorithm Using Cooperative Optimization. Acta Automatica Sinica 35(5), 469–477 (2009)
6. Yang, X.-P., Wang, S.-W., Jia, J.-F.: Color recovery of Dunhuang murals with GrabCut Segmentation and automatic data acquisition. Journal of Lanzhou University of Technology 36(3), 114–117 (2010)
7. Bay, H., Ess, A., Tuytelaars, T., et al.: Speeded-Up Robust Features(SURF). Computer Vision and Image Understanding 110(3), 346–359 (2008)
8. Ge, L., Zhu, Q.-S., Fu, S.-S.: Improved image dense stereo matching algorithm. Journal of Jilin University (Engineering and Technology Edition) 40(1), 212–217 (2010)
9. Li, X., Qin, X.: Structure Scene Reconstruction Based on Quasi-Dense Match Propagation. Journal of Computer-Aided Design & Computer Graphics 23(5), 849–854 (2011)
10. He, L., Liu, R.: An Algorithm for Image Match Propagation. Journal of East China Institute of Technology (Natural Science) 34(2), 379–383 (2011)

Intelligent Scheme of Removing Plug Oil Well Based on Mobile Robot

Xiaojie Tian¹, Yonghong Liu¹, Yulong Xu², Rongju Lin¹, Wei Deng³, and Rui Zhang¹

¹ College of Mechanical and Electronic Engineering, China University of Petroleum,
Dongying, Shandong 257061, China

² Bohai Drilling Engineering Technology Research Institute, Tianjin 300280, China

³ Engineering Institute of Engineering Corps., PLA University of Science and Technology,
Nanjing 210007, China

{tianxj20050101, xuyulongupc, 250960764, flyrockgod}@163.com,
liuyh@upc.edu.cn, 1440644797@qq.com

Abstract. The mobile robot for removing the plug oil well is developed. And the robot has the merits of the wheel-type and crawling-type mechanism and can adapt to different oil pipe sizes. A well functioning prototype has been designed and manufactured. The mathematical model is established for the supporting mechanism which is one of important components of the robot. The deformation and stress of some components of the supporting mechanism is investigated by using the FEM. The components show excellent performance with little displacement and small stress under working condition. The strength and rigid of supporting mechanism is good enough to ensure the reliability of the whole robot mechanism.

Keywords: Mobile robot, Supporting mechanism, FEM, Displacement and stress.

1 Introduction

Over the last decade, mobile robots have been widely used to carry out manifold tasks such as industrial applications, planetary exploration, rescue operation and medical services. In the oil and gas field, there are a lot of pipes that need to be detected and rescued, which promote the development of the mobile robots.

According to a locomotive mechanism to achieve the desired mobility, mobile robots may be split into following categories: leg-type, track-type and wheel-type mobile robots. While the leg-type mobile robot ensures the most superior adaptability to all kinds of environments, its mechanism is quite complicated because active control algorithms equipped with additional actuators and sensors are required to steadily maintain its balance, which inevitably leads to slow movement and poor energy efficiency [1, 2]. The track-type mobile robot provides acceptable mobility on an off-road environment by virtue of its inherently stable mechanism. However, the excessive friction is lost during changing a direction, which also results in poor energy efficiency [3]. Compared to other alternatives, the wheel-type mobile robot

can be developed in the simplest configuration. Therefore the fast movement as well as good energy efficiency is guaranteed without any complicated control strategy. However, its adaptability to an environment does not seem to be sufficiently good and its mobility is restricted depending on both the type and the size of encountered obstacle [4].

Therefore, it is not surprising that high mobility on various environments have been a primary factor among others when evaluation the performance of the mobile robot. Li Peng et al. [5] proposed an adaptive mobile robot which had the adaptability to the change of pipe diameters. Joshi et al. [6] designed a spherical mobile robot, rolling on a plane with the help of two internal rotors. Kim et al. [7] presented an optimal design of a wheel-type mobile robot having high mobile stability as well as excellent adaptability while climbing stairs. Aracil et al. [8] proposed the parallel robots for autonomous climbing along tubular structures and studied the dynamics of some different configurations.

2 Structure Principle of the Robot

2.1 The Whole Mobile Robot System

The whole mobile robot system for removing the plug oil well is shown in Fig.1 (a). As shown in Fig.1 (b), the mobile mechanism is mainly composed of micro-step walking mechanism, revolving measuring mechanism, and EDM removing plug mechanism. When the pipe is plugged, the moving robot is tripped into the oil pipe by the drawworks. Once the robot is in the designated position, the drawworks will stop working and the micro-step walking mechanisms will start moving. While the robot crawls along the inner surface of the pipe, the revolving measuring mechanism rotates to detect the perforating position according to the sensors. When the perforating position has been detected, the micro-step walking mechanism will stop the movement and the EDM removing plug mechanism will remove the plugged objects.

2.2 Micro-step Walking Mechanism

The micro-step walking mechanism is the main member of the mobile robot. It can enable the robot walk and stop in any position of the vertical oil pipe. Also it can ensure the robot centring in the pipe and adaptive to different diameters of the pipe. The micro-step walking mechanism contains two sets of adaptive guiding mechanism, supporting mechanism and electric telescopic rod. Based on the principle of slider-crank mechanism, the adaptive guiding mechanism has four cranks distributed for 90° that are opened by the slider pushing at the effect of the pretightening force of spring. The tension wheels are installed on the adaptive guiding mechanism to reduce the friction force and help the robot tripped into the oil pipe smoothly. The supporting mechanism is driven by the step-motor. When the supporting legs of the supporting mechanism is compressed on to the surface of the pipe, the friction force between supporting legs and pipe is large enough to ensure the robot hovering steadily for a long time. The electric telescopic rod can control the distance per step through controlling its telescopic direction and turn-on time.

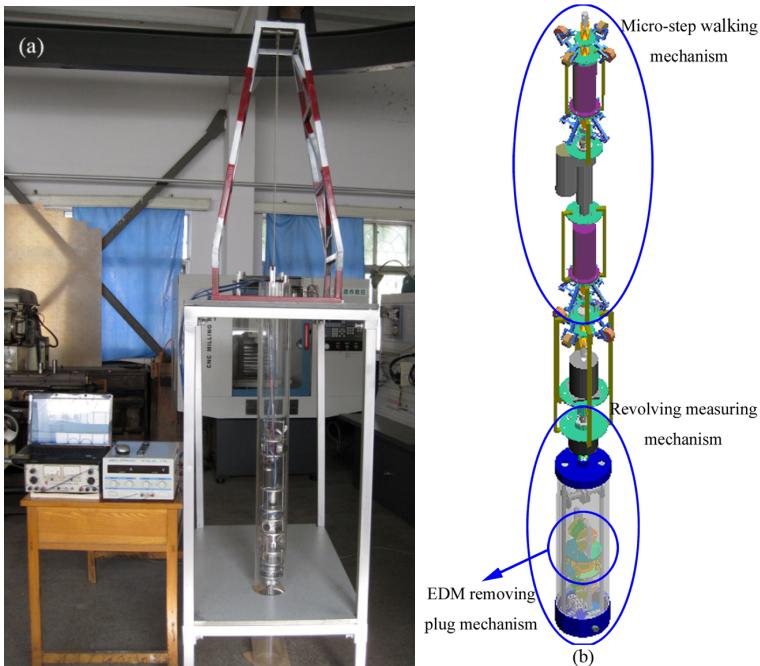


Fig. 1. Schematic diagram of the mobile robot

2.3 Revolving Measuring G Mechanism

The revolving measuring mechanism is responsible for detecting the perforating location in the oil well and can revolve 360° in the pipe, which makes the measuring sensor detect the circumferential surface of the pipe. Moreover the lower part of revolving measuring mechanism is attached with the measuring sensor and the EDM removing plug mechanism distributed symmetrically. Once the perforating location is detected, the revolving measuring mechanism will rotate 180° and the EDM removing plug mechanism is in alignment with the perforating location exactly. So the removing plug work can be carried out. The revolving measuring mechanism should have higher positioning accuracy to ascertain the location of perforating. The measuring work is mainly depending on the electric eddy current sensor which is a non-contacting sensor and produces the output signals according to the eddy current.

3 Mechanical Model for the Supporting Mechanism

3.1 Mathematical Modeling

It is worthwhile to consider the static analysis on the robot mechanism so as to meet the requirement of the strength and rigid of the mechanism. The supporting

mechanism is the most important component in the whole robot mechanism and endures the gravity, the supporting force of the pipe wall and the friction force. In order to analyze the interaction forces between the supporting mechanism and the pipe wall during the EDM removing plug mechanism working condition, we established the mathematical model. Considering the symmetry of the mechanism, the mathematical model is simplified as shown in Fig. 2(a). The supporting leg could be simplified to the two-force bar [9]. The force on the pin A that is contacted with the pipe wall is analyzed in Fig. 2(b). According to the principle of force balance, the force can be expressed as

$$\sum F(0) = 0 \quad (1)$$

$$F_1 = F_0 / 2 \sin(\theta_2) \quad (2)$$

$$F_2 \times L \times \sin(\theta_1) - F_1 \times \sin(\theta_1 + \theta_2) \times L - f \times F_2 \times \cos(\theta_1) = 0 \quad (3)$$

Where F is the resultant force; F_0 is the force produced by the spring on the lower supporting plate; F_1 is the supporting force of the supporting leg; F_2 is the force on the supporting mechanism by the pipe wall, L is the length of the upper supporting leg, f is the friction coefficient between pipe wall and supporting mechanism; θ_1 is the angle of the upper supporting leg to the horizontal line; θ_2 is the angle of the lower supporting leg to the horizontal line.

The force on the supporting mechanism by the pipe wall F_2 can be achieved by Equ. (1), (2), and (3) and expressed as

$$F_2 = \frac{F \sin(\theta_1 + \theta_2)}{2(\sin \theta_1 - f \cos \theta_2)} \quad (4)$$

From Equ. (4), the force on the supporting mechanism by the pipe wall is related to F_2 , θ_1 and has the maximum value only when the sum of θ_1 and θ_2 is the largest and $\sin \theta_1$ is almost equal to $\cos \theta_2$.

3.2 FEM Modelling

FEM software is used to simulate and analyze stress and deformation of the supporting mechanism. The finite element model is established and high precision element SOLID98 element is more suitable for stress and deformation analysis [10, 11]. In addition, the guiding rod that throughput the supporting mechanism (shown in Fig. 1 (b)) is introduced to be analyzed the stress and deformation. It places an important role at the aspect of guiding and supporting the whole robot mechanism. The stainless steel for supporting mechanism and aluminium alloy for guiding rod are considered for simulation. The boundary conditions are fixed on the models and static analyses were performed in sequence in order to obtain the stress and deformation of the components.

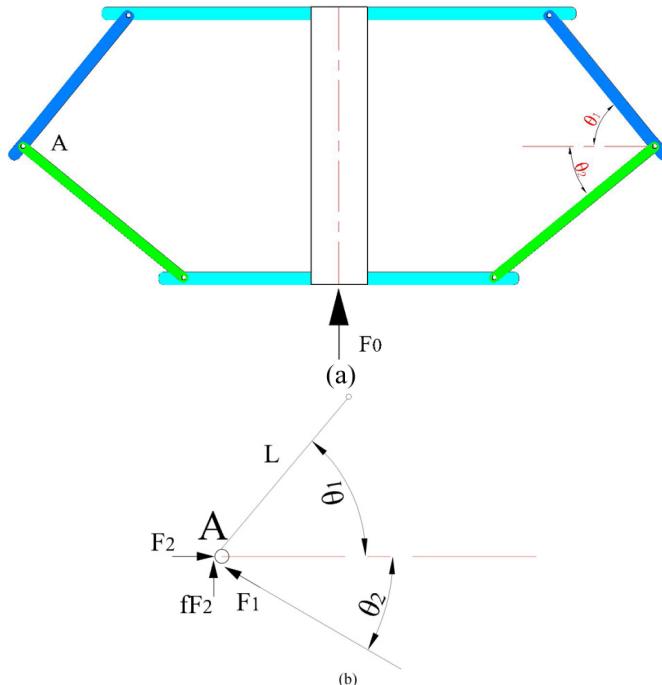


Fig. 2. Mathematical modeling of the supporting mechanism (a) simplified model of the supporting mechanism; (b) mechanical analysis of connecting pin A

4 Results and Discussion

4.1 Displacement and Stress of the Upper Supporting Plate

The displacement of the upper supporting plate is shown in Fig. 3 (a). Under the effect of electromagnetic force and gravity, the maximal deformation emerges at the connecting points of pins. This is mainly owing to the weight of the whole mechanism

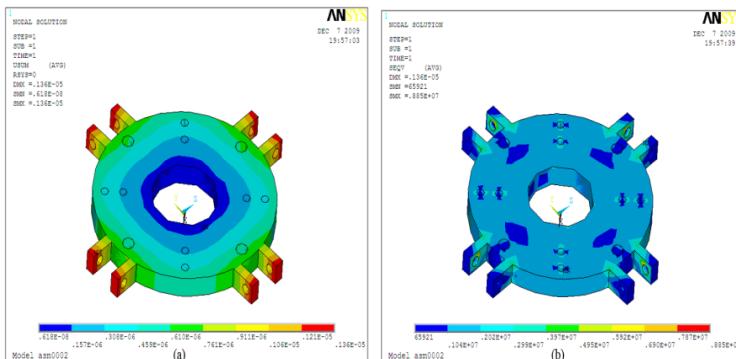


Fig. 3. Simulation results of the upper supporting plate (a) displacement (b) stress

that ultimately caused the connecting points buckled. The stress of the upper supporting plate is shown in Fig. 3 (b). It can be seen that the stress value of the upper supporting plate is only 8.85Mpa, which is much smaller than the allowable stress of material. Therefore the thickness of the upper plate can be reduced appropriately.

4.2 Displacement and Stress of the Supporting Arm

As shown in Fig. 4 (a), the maximal displacement of the supporting arm emerges at the connecting joint with the lower supporting plate. The supporting arm could be regarded as a two force bar whose force is in the direction of its application. Therefore the deformation of supporting arm is in the direction of its application. The supporting arm is the mainly forcing component and the stress is shown in Fig. 4 (b). The stress is a little greater due to the effect of the electromagnetic force on the lower plate. And the greatest value is 48.8Mpa which also meets the strength requirement.

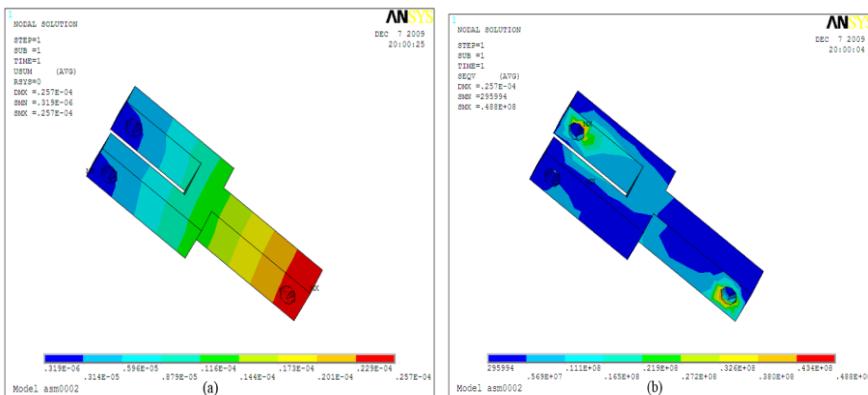


Fig. 4. Simulation results of the supporting arm (a) displacement (b) stress

4.3 Displacement and Stress of the Lower Supporting Plate

The maximal deformation of the lower supporting plate emerged in the centre as shown in Fig. 5 (a), which is differently from the upper supporting plate. This is because that the electromagnetic force is directly put on the middle of the lower plate. The maximal stress value of the lower supporting plate is 56.1Mpa as shown in Fig. 5 (b). Also the stress concentration appears at the sharp angle.

4.4 Displacement and Stress of the Guiding Rod

The deformation of the guiding rod is shown in Fig. 6 (a). The guiding rod mainly bears a tensile force and has the maximal deformation of 0.0021mm. And the maximal stress is 8.96Mpa as shown in Fig. 6 (b), which is far smaller than the allowable stress of aluminium alloy. In one word the guiding rod can meet the requirements of strength and rigid.

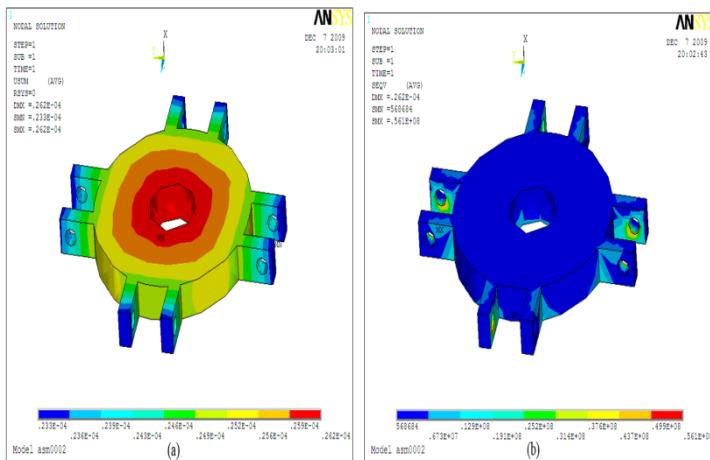


Fig. 5. Simulation results of the lower supporting plate (a) displacement (b) stress

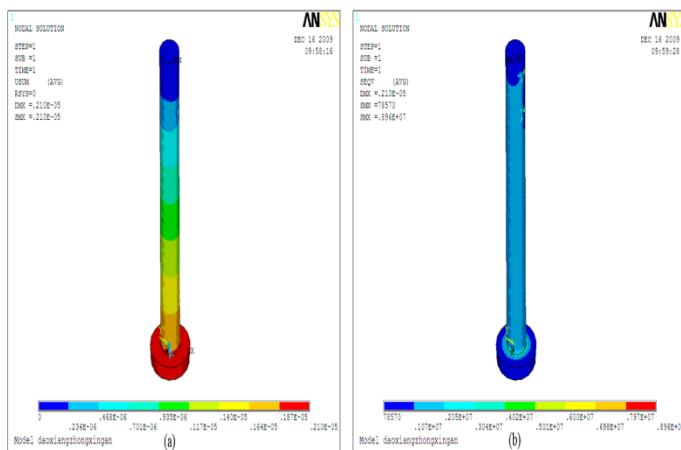


Fig. 6. Simulation results of the guiding rod (a) displacement (b) stress

Acknowledgments. The authors wish to acknowledge the financial support of National High-Technology Research and Development Program of China (No.2007AA09A101), National Natural Science Foundation of China (No.50874115), Taishan Scholar project of Shandong Province (TS20110823), Science and Technology Development Project of Shandong Province (2011GHY11520) and Fundamental Research Funds for the Central Universities (11CX04031A).

References

1. Wu, J., Wang, J., You, Z.: An overview of dynamic parameter identification of robots. *Rob. Comput. Integr. Manuf.* 26, 414–419 (2010)
2. Chernousko, F.L.: Modelling of snake-like locomotion. *Autom. Syst.* 164, 415–434 (2005)
3. Wang, T., Chevallereau, C.: Stability analysis and time-varying walking control for an under-actuated planar biped robot. *Rob. Autom. Syst.* 59, 444–456 (2011)
4. Siggwart, R., Lamon, P., Estier, T., Lauria, M., Piquet, R.: Innovative design for wheeled locomotion in rough terrain. *Rob. Autom. Syst.* 40, 151–162 (2002)
5. Li, P., Ma, S., Li, B., Wang, Y.: Design and motion analysis of an in-pipe robot with adaptability to pipe diameters. *Jixie Gongcheng Xuebao* 45, 154–161 (2009) (in Chinese)
6. Joshi, V.A., Banavar, R.N., Hippalgaonkar, R.: Design and analysis of a spherical mobile robot. *Mech. Mach. Theory* 45, 130–136 (2010)
7. Kim, D., Heeseung, H., Hwa, S.K., Jongwon, K.: Optimal design and kinetic analysis of a stair-climbing mobile robot with rocker-bogie mechanism. *Mech. Mach. Theory* 50, 90–108 (2012)
8. Aracil, R., Saltarén, R., Reinoso, O.: Parallel robots for autonomous climbing along tubular structures. *Rob. Autom. Syst.* 42, 125–134 (2003)
9. Li, Z.Q., Ma, S.G., Li, B., Wang, M.H., Wang, Y.C.: Analysis of the constraint relation between ground and selfadaptive mobile mechanism of a transformable wheel-track robot. *Sci. China Technol. Sci.* 54, 610–624 (2011)
10. Shams, S., Lee, J.Y., Han, C.: Compact and lightweight optical torque sensor for robots with increased range. *Sens. Actuators A Phys.* 173, 81–89 (2012)
11. Zou, J.-T., Chianq, F.-C., Su, K.: The study of path error for an omnidirectional home care mobile robot. *Artif. Life Rob.* 16, 127–131 (2011)

A Novel Meta-heuristic for the Multi-depot Vehicle Routing Problem

Jianping Luo, Xia Li, and Min-Rong Chen

College of Information Engineering, Shenzhen University Shenzhen, 518060, China

Abstract. Shuffled Frog Leaping Algorithm (SFLA), which was presented by Eusuff and Lansey, is a novel meta-heuristic optimization. In this study, an improved SFLA (ISFLA) which improves the leaping rule of frog by enlarging the leaping step size and adding leaping inertia component in the view of social behavior is proposed to solve the Multi-Depot Vehicle Routing Problem (MDVRP). Considering the MDVRP is combinatorial optimization problem, in which every dimension of solution is discrete, we present a novel real number encoding method of ISFLA for MDVRP. At the same time, the Extremal Optimization (EO) is introduced to ISFLA to further improve the local search ability. The results of experiments have shown the proposed algorithm possesses outstanding performance to solve the MDVRP.

Keywords: Integer programming, Combinatorial optimization, Vehicle routing problem, Shuffled frog leaping algorithm.

1 Introduction

The Vehicle Routing Problem (VRP), which was first proposed by Dantzig and Ramser (1959), has been researched extensively because it is widely applicable to many real-world optimization problems and is related to the logistics distribution problem. The Multi-Depot Vehicle Routing Problem (MDVRP) is also an NP-hard problem for simultaneously determining the routes for several vehicles from more than one depot to a set of customers, and then returning to the same depot without exceeding the capacity constraint of each vehicle. Due to the complexity of the problem, solving the MDVRP to optimality is extremely time consuming. To solve the problem efficiently, almost all previous literatures (e.g., Wren et al., 1972; Crevier et al., 2007; Ghoseiri et al., Liu et al., 2010; Mirabi et al., 2010) preferred intelligent optimization methods to exact algorithms. There are two common steps among these proposed methodologies. First, the MDVRP was decomposed, and then the sub-problems were solved sequentially and iteratively. Second, the heuristic methods consisted of two mechanisms: construction and improvement. The first mechanism generated initial feasible solutions, whereas the second mechanism modified the existing solutions to yield better results. However, it is found that very few researchers have applied the optimization algorithms based on swarm intelligence to solve the MDVRP. This is our primary motivation for writing this paper. We present

an improved Shuffled Frog Leaping Algorithm (ISFLA) to solve the MDVRP in this paper.

SFLA which was developed by Eusuff and Lansey (2003) is a meta-heuristic optimization method inspired from the memetic evolution of a group of frogs when seeking for food. In this algorithm evolution of memes is driven by exchange of information among the interactive individuals. It combines the advantages of the genetic-based MA and the social behavior-based Particle Swarm Optimization (PSO) algorithm. SFLA has been applied successfully to many fields and obtains good performance.

This paper is organized as follows: Section 2 describes the hierarchy of decisions in the MDVRP. In Section 3, the SFLA and the improved SFLA is proposed to solve the MDVRP. Experimental evaluations and result discussions are shown in Section 4. Finally, we draw the conclusions.

2 Multi-depot Vehicle Routing Problem

Consider a distribution network in which one product is shipped from some depots to a set of customers. The MDVRP can be defined on multi-graph $G\{C, D, E\}$, where C is customer vertex set which includes N customers, D is the depot vertex set which includes M depots. The set $V = \{C \cup D\}$ is the node set, and E is the arc set. $E = \{(i, j) | i, j \in V\} \setminus \{(i, j) | i, j \in D\}$. The demand of the i -th customer is q_i . The distance between point i and point j is c_{ij} . The system has L vehicles, and the $K = \{k_1, k_2, \dots, k_L\}$ is the set of all vehicles. The maximum capacity of vehicle i is Q_i ($i = 1, 2, \dots, L$). We use $K(d)$ to denote the vehicle set of the depot d ($d \in D$), use K_d to denote the maximum vehicle number of depot d , and $C(d)$ denotes the customer set of the depot, d ($d \in D$). Each vehicle starts the travel from a depot and has to return to the same depot in completion of services to customers. The objective is to determine a viable delivery sequence of each route which minimizes the delivery distance or time spent in serving all customers.

Decision variables:

$$x_{ijk} = \begin{cases} 1 & \text{if point } i \text{ immediately precedes point } j \text{ on route } k; \\ 0 & \text{otherwise} \end{cases}$$

$$y_{kd} = \begin{cases} 1 & \text{if vehicle } k \text{ is allocated to depot } d; \\ 0 & \text{otherwise} \end{cases}$$

The MDVRP can then be stated as follows:

$$Z = \min \sum_{d \in D} \sum_{k \in K(d)} \sum_{i \in (C(d) \cup \{d\})} \sum_{j \in (C(d) \cup \{d\})} c_{ij} x_{ijk} \quad (1)$$

S.t.

$$\sum_{k \in K} \sum_{j \in C} x_{djk} \leq K_d \quad \forall d \in D \quad (2)$$

$$\sum_{d \in D} K_d \leq L \quad (3)$$

$$\sum_{d \in D} y_{kd} = 1 \quad \forall k \in K \quad (4)$$

$$\sum_{k \in K} \sum_{j \in C} x_{ijk} = 1 \quad \forall i \in C \quad (5)$$

$$\sum_{i \in C \setminus p} x_{ipk} = \sum_{j \in C \setminus p} x_{pjk} \quad \forall p \in C, \forall k \in K \quad (6)$$

$$\sum_{j \in C} q_j \sum_{i \in V} x_{ijk} \leq Q_k \quad \forall k \in K \quad (7)$$

$$\sum_{j \in C} x_{djk} y_{kd} = 1 \quad \forall k \in K, \forall d \in D \quad (8)$$

$$\sum_{i \in C} x_{idk} y_{kd} = 1 \quad \forall k \in K, \forall d \in D \quad (9)$$

In the above, Eq.(1) which is the objective function is to minimize the total cost (distance or time) by all vehicles. Eqs. (2) and (3) require that the number of vehicles which are arranged is less than the maximum number of vehicles. Eq.(4) requires that each customer to be assigned to a single route. Eqs.(5) and (6) assure that each route can be served at most once. Eq.(7) is the capacity constraint set for vehicles. Eqs.(8) and (9) require that every route is circular path.

3 Improved Shuffled Frog Leaping Algorithm for MDVRP

3.1 Shuffled Frog Leaping Algorithm

SFLA is a intelligent optimization method that inspirited the memetic evolution of a group of frogs in pond when searching for the point that has the maximum amount of food. It is described in details as follows. First, an initial population of F frogs is created randomly. For the d -dimensional problem, the position of the ‘ith’ frog is represented as $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$. Afterwards, the frogs are sorted in a descending order according to their fitness. Then, the F frogs are separated into m memplexes according to their fitness ordering, each containing n frogs (i.e. $F = mn$), in such a way that the first frog goes to the first memplex, the second frog goes to the second memplex, the m th frog goes to the m th memplex, and the $(m+1)$ th frog goes back to the first memplex, etc. The main work of algorithm is to update the position of the worst performance frog through iterative operation in each

memeplex. Its position is improved by learning from the best frog of the memeplex or the population and position of its own. In each memeplex, the new position of the worst frog is updated according to Eqs. (10) and (11):

$$D = r.(x_s - x_w(k)) \quad (10)$$

$$x_w(k+1) = x_w(k) + D, \quad \|D_{\min}\| \leq \|D\| \leq \|D_{\max}\| \quad (11)$$

where $x_w(k)$ and x_s are the worst frog position and the best frog position respectively in the memeplex, r is a random number in range[0,1], k is the iteration number of the memeplex, and $\|D_{\min}\|$ and $\|D_{\max}\|$ are the maximum and minimum allowed change in a frog's position respectively. If the new position of the worst frog is better than before, it replaces the worst frog's position. Otherwise, the calculations in Eqs.(10) and (11) are repeated but with respect to the global best frog (i.e. x_b replaces x_s). If this process still can't obtain the better performance, the position is randomly generated for the worst frog. The iteration continues for a predefined number of memetic evolutionary times within each memeplex, and then the whole population is mixed together in the shuffling process. The local evolution and global shuffling continue until convergence criteria are satisfied.

3.2 Improved Shuffled Frog Leaping Algorithm (ISFLA)

Considering a group of frogs leaping in a swamp, the frog whose position has the minimum amount of available food will tend to jump toward best frog's position where has more food through communicating memes(information) with each other. It is reasonable that the area around this best position maybe exists some position where possesses more food than current best position. This position might locate beyond the range between current worst position and best position. The frog leaping rule in the SFLA is inspired from this social imitation. However, on the basis of the original frog leaping rule presented above, the leaping of the worst frog is restricted in the range between its current position and the best frog's position, and never jump over the position of the best one. This limited and narrow leaping method might slow down the convergence speed, and even cause premature convergence. To enlarge the search range, the formula (10) can be modified as:

$$D = r.c.(x_s - x_w(k)) \quad (12)$$

where c is a constant greater than 1, which presents the searching scale of every leaping step for one frog. The value of c can't be too large, otherwise, the searching range is too wide and the search method will be developed into random search which is slow in convergence speed.

On the other hand, According to the view of social behavior, while leaping for finding food, every frog might has a movement inertia which represents self confidence for current movement status for itself. Every frog adjusts leaping step and determines the leaping direction not only by the referring to the best frog's position,

but also in view of the self movement inertia. This idea leads to a new leaping rule to replace formula (12). The new leaping rule is expressed as:

$$D_w(k+1) = w.D_w(k) + r.c.(x_s - x_w(k)) \quad (13)$$

$$x_w(k+1) = x_w(k) + D_w(k+1), \quad \|D_{Min}\| \leq \|D\| \leq \|D_{max}\| \quad (14)$$

Here $D_w(k)$ can be seen as the leaping velocity of the worst frog in the k th iteration.

It is noticeable that leaping inertia component of new leaping rule might lightly slow down the convergence speed for it is more difficult to change the movement of frog, but it will help frog to move more stably and improve the global search ability.

Step1:

Divide X into group, generate M sets,

denoted by G_1, \dots, G_M , where

$$G_1 = \{(x_i, i) \mid 0 \leq x_i < K_1\},$$

$$G_2 = \{(x_i, i) \mid 1 \leq x_i < K_2\}, \dots$$

$$G_M = \{(x_i, i) \mid K_{M-1} \leq x_i < K_M\};$$

Step2:

Divide set G_m ($1 \leq m \leq M$) into group,

generate K_m sets.

When $m = 1$

$$J_{1,1} = \{(x_i, i) \mid 0 \leq x_i < 1\},$$

$$J_{1,2} = \{(x_i, i) \mid 1 \leq x_i < 2\},$$

...

$$J_{1,K_1} = \{(x_i, i) \mid K_1 - 1 \leq x_i < K_1\};$$

When $1 < m \leq M$,

$$J_{m,1} = \{(x_i, i) \mid K_1 + \dots + K_{m-1} \leq x_i < K_1 + \dots + K_{m-1} + 1\},$$

$$J_{m,2} =$$

$$\{(x_i, i) \mid K_1 + \dots + K_{m-1} + 1 \leq x_i < K_1 + \dots + K_{m-1} + 2\},$$

...

$$J_{m,K_m} = \{(x_i, i) \mid K_1 + \dots + K_{m-1} + K_m - 1$$

$$\leq x_i < K_1 + \dots + K_{m-1} + K_m\};$$

In the above, every set denotes a vehicle route.

Step3:

Sort the elements of set $J_{m,k}$ according to the x_i in ascending order, and the second dimension value of

Fig. 1. The decoding process of a MDVRP solution

3.3 ISFLA for MDVRP

The original model of ISFLA is adapted to solve the continuous optimization problems, but it is hard to handle the MDVRP directly. The MDVRP is combinatorial optimization problem, in which every dimension of solution is discrete. So we ought to encode the solution of problem and make it suitable for ISFLA to calculate. In this paper, we use a novel real number encoding method of ISFLA for MDVRP.

For the MDVRP including M depots and N customers, the data structure of the SFLA solution can be described as : $X = (x_1, x_2, \dots, x_N)$, $x_i \in [0, K_1 + K_2 + \dots + K_M]$, where K_m ($m \in 1, \dots, M$) is the maximum number of the depot m . The data structure includes complete information of a MDVRP solution. Its decoding process is shown in Fig. 1.

The every element of solution is real number in this encoding method, so the iteration updating Eqs. (13) and (14) can be used directly to solve the MDVRP.

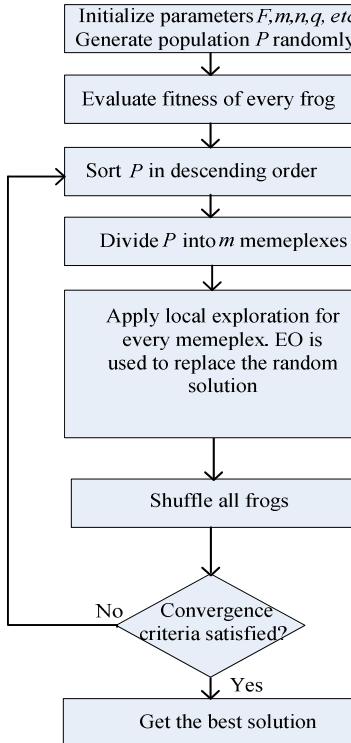


Fig. 2. The flow chart of the ISFLA-EO

3.4 ISFLA with Extremal Optimization (EO)

Extremal Optimization (EO), proposed by Boettcher and Percus (1999), is a general-purpose local search optimization approach. We have proved that the SFLA-EO algorithm is very effective and superior to solve continuous global optimization problems (2009). In this work, we take the EO as the local search method to solve the MDVRP. Considering that EO has strong local-search ability, in order to further improve the local search ability of ISFLA, we present a novel hybrid ISFLA-EO algorithm which makes full use of the exploration ability of ISFLA and the exploitation ability of EO. The random solution of ISFLA promotes the randomness and the diversity of algorithms, but it weakens the convergence of algorithm. To introduce EO into ISFLA, we use EO process to replace the random solution generation process. It means that all frogs will execute the EO local search at least one time if the EO operation is used in ISFLA. The ISFLA-EO can be summarized as Fig.2.

4 Experiments and Results

Note that our algorithms were run on the hardware (i.e., Intel Pentium 4 with 2.8 GHz CPU and 512M memory) and software (i.e., VC++) platform. The parameters setting of ISFLA-EO are performed: $m = 20, n = 20$, and the generation number for the memeplex is equal to 10. These values are suitable to get good solutions in terms of the quality of the solution and the processing time in accordance with our experiments observation. In experiments, the solution stops when the following criterion is satisfied: the objective function value is not improved in 60 successive shuffling iterations.

Tables 1 presents the results of two benchmark heuristics and the proposed multi-phase/single-phase algorithm on 21 MDVRP instances, respectively. The test data set used for the comparison come from Laporte, Cordeau, and Mercier (2001). These instances are available at (<http://www.hec.ca/chairedistributique/data>). The first three columns of Table 1 describe the benchmark instances with instance name, the number of customers which have to be served and the number of depots denoted by Name, n and d, respectively. In Table 1, we present the best, average result and average runtime over 10 runs for all the randomized methods, the time is recorded in minutes. The fitness of a solution is recorded as the total distance traveled by the vehicles. The ALNS framework presented by Crevier et al. (2007) is an extension of the large neighborhood search framework by Shaw [Using constraint programming and local search methods to solve vehicle routing problems. In: CP-98, Fourth international conference on principles and practice of constraint programming, Lecture notes in computer science, vol. 1520, 1998. p.17–31] with an adaptive layer (hardware: 3 GHz Pentium 4). This layer adaptively chooses among a number of insertion and removal heuristics to intensify and diversify the search. Villegas et al. (2010) propose GRASP/VND which contains some meta-heuristics based on variable neighborhood descent (VND), greedy randomized adaptive search procedures (GRASP) and evolutionary local search (ELS) to solve MDVRP (hardware: 3.4GHz Pentium D).

Note that GRASP/VND didn't solve the instances of capacity constrained and route duration constrained MDVRP, which are in the last 10 rows, i.e., from pr01 to pr10, we use the sign “-” to indicate it in Table 1.

From these results some conclusions can be drawn: (1) The ISFLA-EO algorithm presented in this paper got 20 out of 21 best solutions. It proves that our algorithm isn't apt to trap to the local optima, and has excellent ability to escape extremal points. (2) On the convergence speed of algorithms, it is obvious that ISFLA-EO is very fast. The hardware environments of three algorithms are about the same. The runtime of ISFLA-EO spent on finding global optima is much less than that of GRASP/VND algorithms. For instance, in contrast to that GRASP/VND needed 3.55 minutes to get the global optimum in p07, ISFLA-EO spent only 0.96 minute to finish the search. It should be noted that the runtime of ISFLA-EO is less than that of ALNS, when the number of customers is not too large. However, with increasing the customers number, the runtime raise and it is close to the ALNS's runtime. This is because SFLA has many individuals, which lead iteration of ISFLA to be a time-consuming process. (3) The table also shows that the ISFLA-EO is quite stable as the average solutions are very close to the best solutions and the average deviations never surpass 1% in all instances.

Table 1. Comparisons of some algorithms applied to the MDVRP

Instance	ALNS				GRASP/VND				ISFLA-EO			
	Name	n	d	AVE	BEST	TIME	AVE	BEST	TIME	AVE	BEST	TIME
p01	50	4	576.87	576.87	0.48	576.87	576.87	0.89	576.87	576.87	0.25	
p02	50	4	473.53	473.53	0.47	473.53	473.53	1.12	473.53	473.53	0.31	
p03	75	5	641.19	641.19	1.07	643.44	641.19	2.09	641.19	641.19	0.82	
p04	100	2	1006.09	1001.04	1.47	1008.1	1003.62	3.5	1004.30	1001.04	0.89	
p05	100	2	752.34	751.26	2	752.54	751.15	3.13	751.62	750.03	1.11	
p06	100	3	883.01	876.7	1.55	884.42	880.69	3.51	880.59	876.7	1.05	
p07	100	4	889.36	881.97	1.47	892.59	888.65	3.55	890.05	881.97	0.96	
p12	80	2	1319.13	1318.95	1.25	1318.95	1318.95	2.2	1318.99	1318.95	0.82	
p15	160	4	2519.64	2505.42	4.22	2508.05	2505.42	9.3	2509.22	2505.42	4.19	
p18	240	6	3736.53	3702.85	6.98	3737.64	3702.82	22.72	3734.21	3702.55	7.50	
p21	360	9	5501.58	5474.84	9.7	5522.02	5490.55	52.96	5502.15	5474.84	13.12	
pr01	48	4	861.32	861.32	0.5	-	-	-	861.32	861.32	0.31	
pr02	96	4	1308.17	1307.34	1.72	-	-	-	1307.34	1307.34	1.12	
pr03	144	4	1810.66	1806.6	3.57	-	-	-	1808.92	1806.6	2.88	
pr04	192	4	2073.16	2060.93	4.93	-	-	-	2066.58	2058.83	5.21	
pr05	240	4	2350.31	2337.84	6.2	-	-	-	2347.12	2336.57	7.12	
pr06	288	4	2695.74	2687.6	7.75	-	-	-	2694.72	2685.24	8.40	
pr07	72	6	1089.56	1089.56	0.97	-	-	-	1089.56	1089.56	0.55	
pr08	144	6	1675.74	1664.85	3.45	-	-	-	1666.21	1664.85	2.86	
pr09	216	6	2144.84	2136.42	5.83	-	-	-	2142.47	2136.42	5.68	
pr10	288	6	2905.43	2889.82	7.58	-	-	-	2912.74	2890.16	8.12	

5 Conclusion

In this paper, we present an improved Shuffled Frog Leaping Algorithm (ISFLA) based on real-coded patterns to solve the Multi-depots Vehicle Routing Problem (MDVRP). A hybrid Extremal Optimization (ISFLA-EO) which introduces EO local search method to ISFLA is also introduced in order to further improve the local search ability of ISFLA. The results of experiments have shown the proposed algorithm possesses outstanding performance to solve the MDVRP.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grant No.60772148 and No.61005049.

References

1. Boettcher, S., Percus, A.G.: Extremal Optimization: Methods derived from Co-Evolution. In: Proceedings of the Genetic and Evolutionary Computation Conference, New York, USA, pp. 101–106 (1999)
2. Crevier, B., Cordeau, J., Laporte, G.: The multi-depot vehicle routing problem with inter-depot routes. European Journal of Operational Research 176(2), 756–773 (2007)
3. Dantzig, G., Ramser, J.: The truck dispatching problem. Management Science 6(1), 80–91 (1959)
4. Eusuff, M., Lansey, K.: Optimization of Water Distribution Network Design Using the Shuffled Frog Leaping Algorithm. Journal of Water Resources Planning and Management 129(3), 10–25 (2003)
5. Ghoseiri, K., Ghannadpour, S.F.: A hybrid genetic algorithm for multi-depot homogenous locomotive assignment with time windows. Applied Soft Computing 10(1), 53–65 (2010)
6. Liu, R., Jiang, Z.B., Fung, R.Y.K., Chen, F., Liu, X.: Two-phase heuristic algorithms for full truckloads multi-depot capacitated vehicle routing problem in carrier collaboration. Computers & Operations Research 37(11), 950–959 (2010)
7. Luo, J.P., Chen, M.R., Li, X.: A Novel Hybrid Algorithm for Global Optimization Based on EO and SFLA. In: 2009 4th IEEE Conference on Industrial Electronics and Applications, ICIEA, pp. 1935–1939 (2009)
8. Mirabi, M., Fatemi Ghomi, S.M.T., Jolai, F.: Efficient stochastic hybrid heuristics for the multi-depot vehicle routing problem. Robotics and Computer-Integrated Manufacturing 26(6), 564–569 (2010)
9. Villegas, J.G., Christian, P.: GRASP/VND and multi-start evolutionary local search for the single truck and trailer routing problem with satellite depots. Engineering Applications of Artificial Intelligence 23(5), 780–794 (2010)
10. Wren, A., Holliday, A.: Computer scheduling of vehicles from one or more depots to a number of delivery points. Operational Research Quarterly 23(3), 333–344 (1972)

An Improved Association Rule Privacy Protection Algorithm Research

Yiwu Chen, Qingshui Li^{*}, and Lingna He

Computer Science and Technology College,
Zhejiang University of Technology Hangzhou, China
mdusa@zjut.edu.cn

Abstract. In view of the current privacy protection mining algorithm existing the defects, such as data protection level is not high, transaction database of rules loss rate is bigger problems, paper proposes an improved association rules privacy protection algorithm, to better solve the problems of the association rules privacy protection. In the privacy of the same rate, improve the privacy information protected extent and reduce the rules loss of the transaction database, to greater extent to privacy protection. The simulation results show that, the improved privacy protection association rule mining algorithm in data protection extent has better privacy than the original algorithm, and generated the loss rate and dissimilarity more lowly.

Keywords: association rules, privacy protection, PPARM, SWTA.

1 Introduction

The privacy protection data mining's concept was put forward unceasingly in the recent several years, The connection rule's data mining most early is raised by R.Agrawal, nowadays already became a most important data mining technology, what it mainly solves is makes the extraction useful project relations or rule in the database to the massive item sets[1]. In 1995 called the first academic conference, the first time as one of the future research direction of data mining content based on privacy protection [2]. Later, on the privacy protection of data mining researchers are more and more paying attention, especially for association rules of the new algorithm and the new theory have been proposed by the researchers. Research of association rule privacy protection technology has a very important theoretical value and practical significance.

2 Privacy Protection Technology

Privacy protection is accompanied by data and sensitive rules and put forward. In the field of statistics, privacy protection issues most attention by researchers. With the rapid development of the database technology, all trades and professions have

^{*} Corresponding author.

accumulated a large amount of data. How to extract for decision-making valuable knowledge from these data, at the same time don't leak out sensitive data become very necessary [3]. the main research direction of privacy protection is show in Table 1.

Table 1. Privacy protection major research directions table

Research direction	Sample
General privacy protection technology	Perturbatio, Randomization, Swapping, Encryption
Data Mining Oriented privacy protection technology	Association Rule Mining, Classification Clustering
Data distribution technology based on privacy protection	K-annoymity, L-diversity, M-Invariance closeners
Privacy protection algorithm	Anonymized publication, Anonymization with high utility

The privacy preservation technology may divide into the central data the privacy protection and the distributional data privacy protection. The vertical distribution is mainly refers to the data to distribute according to the attribute in many organizations [4]. The privacy maintains the technology is mainly divided three kinds:

(1) Privacy maintains technology based on the heuristic. (2) Privacy maintains technology based on the information security theory. (3) Privacy maintains technology based on the restructuring privacy. Generally, privacy protection excavation algorithm based on the connection rule classification sees Table 2.

Table 2. Privacy Protection Based on Association Rules Operator Classification

Privacy protection technology	data change way	Data mining algorithms
Heuristic	Slide window method Randomly modify some value1the data is 0 Add random number	Association rules
Reconstruction type	Random changes part of data	Association rules

3 Association Rules Algorithm

The association rule excavation generally speaking is refers to from a great quantity data discovered that the interesting incidence relation, namely distinguishes the attribute value collection which from the data set appears frequently, is also called the frequent item set. Then on this basis, from the discovered frequent set to find association rules method, This reveals some dependence relation between data items, according to the last generation rules, which can be found relations from the object [4]. Apriori algorithm is a kind of mining Boolean association rule frequent influential classical algorithm. the basic framework of the Apriori algorithm are described below:

Algorithm description: using layer-by-layer iterative method based on the candidate generation to find frequent item sets

Input: D: initial database minimum support min_sup threshold

Output: L: D frequent item sets

- (1) L1=find_frequent_1-itemsets (D);
- (2) For(k=2;Lk-1 ≠ ∅ ;k++){
- (3) Ck=Apriori_gen(Lk-1);
- (4) for each affairs t∈ D
- (5) Ct= subset (Ct);
- (6) for each Candidate c∈ Ct
- (7) C.count++;
- (8)}
- (9) Lk=(C∈ Ck |C.conut≥ min_sup)
- (10) }
- (11) Return= $U_k L_k$;

The outstanding characteristic of Apriori algorithm is use the K-1 trip from scanning strengths set L_{k-1} to generate candidate set C_k , It passes through (k-1) strengths set L_{k-1} afterlife as parameters of the function Apriori_gen (L) to realize, the function returns K strengths set L superset, i.e. candidate C.

4 Classic Privacy Protection Algorithm

privacy protection based on association rules data mining , through the hide sensitive data or sensitive rules to achieve, which is through the selective modify the initial data, after the data cleaning, to achieve sensitive rules of the hidden [5].

4.1 PPARM Algorithm

Makes $I=\{ i_1, i_2, \dots, i_m \}$ is business database DB item set, $T=\{ T_1, T_2, \dots, T_n \}$ is all business's set in the storehouse. Each business T_j contains the item set is I subset. When connection rule is extraction, contains many sets to be called the item set. If an item set contains k item, is called the k item set.

The connection rule is the shape like $X \Rightarrow Y$ implication expression, x and y does not intersect item set, namely $X \cap Y = \emptyset$, Connection rule intensity available support and confidence measure. The support description assigns the item set the frequent degree, namely importance. The confidence determined y in contains the frequent degree which in x business appears, namely the item relates. The formula is as follows:

$$\text{sup port}(x \Rightarrow y) = \frac{|x \cup y|}{|N|} \quad (1)$$

$$\text{confidence}(x \Rightarrow y) = \frac{|x \cup y|}{|x|} \quad (2)$$

And, $x \subseteq I$, $y \subseteq I$, N is business database DB contains business's integer; $|A \cup B|$ simultaneously contain x and the y business integer; $|x|$ is in DB contains x business integer. The connection rule from essentially is the conditional probability: When A occurs, B simultaneously appears the probability has in a big way. The connection rule excavation is discovers these supports and the confidence are bigger than define in advance threshold value minsupp and the minconf rule.

The regular support and the confidence formula obtain by the connection rule basic concept, maintains the N sum separately under $|x|$ invariable premise, if suitably reduces $|y|$ value, then may reduce rule $x \Rightarrow y$ the support and the confidence, either the confidence is smaller than until the support minSupp which or minConf assigns, This time, this connection rule no longer satisfies was discovered needs condition, thus achieved hid this sensitive connection rule the goal.

4.2 Performance Evaluation of Privacy Protection

Privacy protection technology needs in the protection privacy of information at the same time, taking into account the practical application value and the computation overhead. In general, usually from the following three aspects of privacy protection technology to measure [6]:

(1) The rate of privacy

This indicator is often used to illustrate the original data privacy protection, privacy rate is higher, the higher the degree of privacy protection.

(2) Rules loss rate

This index is published on the measures of data quality, reflecting the privacy protection technology of the data processed information loss degree. The higher rate of rules lost, lost more information of association rules,, the more data utilization rate is lower.

(3) The algorithm performance

This index generally uses time complexity and space complexity for the algorithm performance measurement. In general, the smaller of the time complexity, the better the performance of the algorithm.

5 Improvement Association Rules Privacy Protection Algorithm

PPARM algorithm is considered sensitive item sets to do less time delete operation, do the update to the database, to hide the purpose of sensitive association rules. But it has not solved to the database the influence of the non sensitive rules. Therefore, also cannot very good solve practical application protection privacy. Improved association rule algorithm of privacy protection -SWTA algorithm can achieve a higher sense of privacy [7].

SWTA Algorithm Description as follows: In Affairs library collection {T1: ABCD, T2:CD,T3:ABCDEF,T4:BDF,T5:ABCD,T6:BCF,T7:ABD,T8:BCDEF} The stipulation smallest support is 40%, the smallest confidence is 60%. Calculation support{A → B}=4/8=50%, confidence{ A → B }=4/4=100% May know by the PPARM algorithm a=1,b=2,k=min(a,b)=1,Only needs to contain B sensitive business to delete 1 time, can cause its confidence to reduce 37.5% If using the classic Apriority algorithm obtain nine association rules, if rules A → B, B → F, E → F is sensitive rules, then the corresponding set of sensitive items SI = {{A, B}, {B, F}, {E, F}}, Algorithms are dealing with sensitive item sets Si={A,B}, At this point pending the collection of sensitive items SI={{B,F},{E,F}},Num(B,{A,B}, SI,T1)=1, Num=(A,{A,B},SI'T1)=0. It can be seen, if removed A from the transaction T₁, not SI 'sensitive items in the impact, if out of the project B, then it will affect SI' in the sensitive item set {B, F}.

In order to minimize the number of database modifications, the algorithm has a large selection out of the value of Num sensitive items, so as to ensure sensitive items hidden while the lower set of items to be dealing with sensitive support. If in the Num Sj the same value of different items, select a lower moving support items. Like this is for the relative reduction to the database the processing number of times. It can realize the relatively few motion data simultaneously to realize many hides the sensitive rule as far as possible, see Table 3.

Table 3. Transaction database

TID	A	B	C	D	E	F
T1	1	1	1	1	0	0
T2	0	0	1	1	0	0
T3	1	1	1	1	1	1
T4	0	1	0	1	0	1
T5	1	1	1	1	0	0
T6	0	1	1	0	0	1
T7	1	1	0	1	0	0
T8	0	1	1	1	1	1

6 The Analysis of Experimental Results

In order to evaluate the usefulness of this algorithm, cooperated and analyses the PPARM algorithm and SWTA algorithm loss rate and dissimilarity. Experimental operating system is Windows XP, database is Microsoft office Access 2003, with VC ++ language implementation. the test pass to many data set to carry on a great deal of experiment verification, recorded PPARM algorithm and SWTA algorithm under the parameter the same circumstance, several experiment data set of movement result, see table 4 and table 5.

Table 4. Time-consuming compared

Item number	Support	confidence	Conceal rate	PPARM	SWTA
10	0.2	0.3	0.1	0.625	0.710
			0.2	0.714	0.826
			0.3	0.813	0.877
			0.4	0.813	0.963
50	0.3	0.4	0.1	5.179	5.362
			0.2	5.289	5.904
			0.3	5.671	6.382
			0.4	5.179	6.616
100	0.4	0.5	0.1	17.434	19.493
			0.2	18.439	22.769
			0.3	20.127	25.785
			0.4	23.576	27.974

Table 5. Loss rate compared

Item number	Support	confidence	Conceal rate	PPARM	SWTA
10	0.2	0.3	0.1	22.058	19.045
			0.2	23.064	20.947
			0.3	24.157	21.069
			0.4	25.869	22.368
50	0.3	0.4	0.1	11.265	8.953
			0.2	12.573	10.101
			0.3	14.468	11.479
			0.4	16.847	14.421
100	0.4	0.5	0.1	17.434	12.984
			0.2	18.439	14.170
			0.3	20.127	16.411
			0.4	23.576	19.188

From the table 4 and the table 5 can analyze, if give to settle same function parameter, to two algorithms, the PPARM algorithm solved the influence to the non-sensitive rule, so consume to take a lot of time more. But more important, Be compared to PPARM algorithms, But more important, Be compared to PPARM algorithms, SWTA mining algorithm creation the rule throw rate much lower. While handling sensitive item, it nicely carried on a choice to the processed sensitive item wanted and passed the reasonable ambulation to the sensitive item and nicely solved the PPARM algorithms produces to rule loss that the lead bigger blemish, Because SWTA algorithms in the process of letting up sensitive item support number to non-sensitive item of the influence did repair, therefore make original database of the quality got an exaltation, rule dissimilitude degree the variety is smaller.

To the same data set, under other parameters and conceal rate same condition, Compare to PPARM algorithm, under usually condition SWTA algorithm be always lower than the rule that the PPARM algorithm generates to throw lose rate. According to record several data of experiments, result such show as Figure 1.

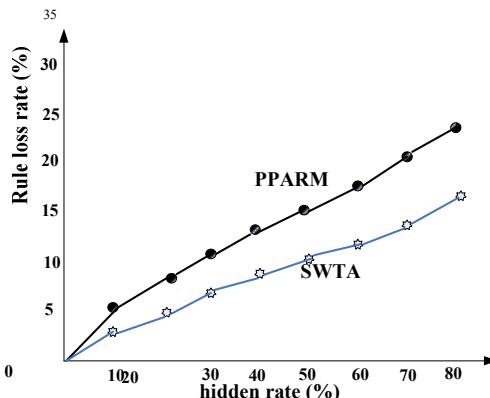


Fig. 1. The relation graph of hidden rates and rule loss rate

7 Summary

This paper summarizes the existing privacy protection technology, analysis of association rules privacy protection mining algorithm, in-depth study of privacy protection association rule mining algorithm (PPARM), In terms of theory, put forward at the same time resolve sensitive rules and non sensitive rules for transaction database the effect of SWTA algorithm, To the same transaction database, in the same hidden rate effectively reduce the loss rate of the original database rules.

References

1. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases, pp. 50–65 (2005)
2. Wang, G., JiangPing: Data mining review. Journal of tongji university (natural science edition) (02) (2004)
3. Zhou, S., Li, F., Tao, Y.: Study on privacy protection for database application. Journal of Computer 32(5), 847–861 (2009)
4. Wu, R.: Data mining algorithm based on association rules. Inner Mongolia Agricultural University (2010)
5. Fang, J.: association rule mining algorithm based on privacy preserving. Zhongshan University, pp. 6–11 (2007)
6. Li, X., Liu, Z.: He privacy protection in data mining. Application Research of Computers 25(12), 3550–3555 (2008)
7. Zhang, R., Zheng, C.: Association rule mining algorithm based on privacy protection. Computer Engineering 35(4), 78–82 (2009)

Multi-step Prediction of Volterra Neural Network for Traffic Flow Based on Chaos Algorithm

Lisheng Yin, Yigang He, Xueping Dong, and Zhaoquan Lu

School of Electrical and Automation Engineering, Hefei University of Technology,
230009, Hefei, China

yls20000@163.com, hyghnu@yahoo.com.cn, {hfdxp, luzhquan}@126.com

Abstract. For the multi-step prediction object of traffic flow chaotic time series, a fast learning algorithms of VNNTF based on chaotic mechanism was proposed. First, combination of chaotic phase space reconstruction properties to traffic flow chaotic time series, method of the truncation order and the truncation items is given, and the VNNTF neural networks model was build by this. Second, based on chaotic learning algorithm, and designed neural network traffic Volterra learning algorithm for fast learning algorithm. Last, a multi-step prediction of traffic flow chaotic time series is researched by VNNTF network model, Volterra prediction filter and the BP neural network based on chaotic algorithm. The results showed that the VNNTF network model predictive performance is better than the Volterra prediction filter and the BP neural network by the simulation results and root-mean-square value.

Keywords: Chaos Theory, Phase Space Reconstruction, Time Series Prediction, VNN Neural Networks, Algorithm.

1 Introduction

The Volterra series is a model for non-linear behavior similar to the Taylor series. It differs from the Taylor series in its ability to capture 'memory' effects. It has the advantages of high precision and clear physical meaning [1]. Traffic flow chaotic time series with the nonlinear behavior of the response and memory function, the Volterra series to become one of the primary means of traffic flow in nonlinear system identification. But the establishment of nonlinear systems on Volterra Series model is very difficult [2-5]. Volterra series has an obvious drawback is that if you want to achieve a satisfactory accuracy may require a considerable number of estimated parameters. The high-level nuclear estimates are facing the greatest difficulties.

The neural network not only has the self-adaptive, parallelism and fault tolerance characteristics, but also has the ability to approximate any nonlinear function. Based on these advantages, the neural network model of the nonlinear system has a very wide range of applications[6-8]. Based on the above considerations, the physical significance of the truncation order of the Volterra series model and the truncated number and the mathematical properties of the minimum embedding dimension and

delay time in traffic flow chaotic time series reconstructed phase space, thus, traffic flow chaotic time series VNNTF network model and the corresponding algorithm has been established[9-12]. VNNIF neural network model,in the traffic flow chaotic time series prediction, obtained good results.

2 Traffic Flow Chaotic Time Series Volterra Model

2.1 The Determination of the Truncation Order on Traffic Flow Chaotic Time Series Volterra Model

Assume that the measured traffic flow chaotic time series is $\{x(t) | (t = 1, 2, 3, \dots)\}$, the input of the nonlinear system is $x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m-1)\tau)$ based on Takens Theorem,where, m is the embedding dimension, τ is the delay time, then the traffic flow chaotic time series phase space reconstruction model with m-order truncation Volterra series model can be characterized as follows:

$$\begin{aligned} x(t'+T) = F(X(t)) &= h_0 + \sum_{l_0=0}^{\infty} h_1(l_0)x(t-l_0\tau) + \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} h_2(l_1, l_2)x(t-l_1\tau)x(t-l_2\tau) + \dots \\ &+ \sum_{l_1=0}^{\infty} \sum_{l_2=0}^{\infty} \dots \sum_{l_p=0}^{\infty} h_m(l_1, l_2, \dots, l_m)x(t-l_1\tau)x(t-l_2\tau)\dots x(t-l_m\tau) \end{aligned} \quad (2)$$

Where, $n, l_m \in R$, $x(t'+T)$ is the output of the nonlinear system; $x(t-l_m\tau)$ is the input of the nonlinear system, $h_m(l_1, l_2, \dots, l_m)$ is the m order Volterra kernel function, $t' = t + (m-1)\tau$, $T (T > 0)$ is the forward prediction step. The total number of items of Volterra series number decreases exponentially growth. In practice, the truncation order is generally the second-order truncation or third order intercept.

2.2 The Determination of the Truncation Items on Traffic Flow Chaotic Time Series Volterra Model

For traffic flow chaotic time series, it is assumed that $x(t)$ and $y(t)$ are the input and output signals of the functional system $f(t, x(t'), t' \leq t)$ in the traffic flow, then:

Traffic flow input signal is a causal relationship: if $t < 0$, then $x(t) = 0$.

Traffic flow functional system $f(t, x(t'), t' \leq t)$ is the limited memory, that is, for $t_0 \rightarrow \infty$, $x(t-t_0)$ has no effect on $y(t)$.

In the prediction of chaos traffic flow chaotic time series, $t' = t + (m-1)\tau$, $T (T > 0)$ is forward prediction step, $x(t'+T)$ represents the output associated with the input signal $x(t)$ and the delay time τ , then

$$\begin{aligned}
x(t'+T) = f(x_{l_1}, x_{l_2}, \dots, x_{N_{l_i}}) &= h_0 + \sum_{l_1=0}^{N_{l_1}-1} h_1(l_1)x(t-l_1\tau) + \sum_{l_1=0}^{N_{l_2}-1} \sum_{l_2=0}^{N_{l_2}-1} h_2(l_1, l_2)x(t-l_1\tau)x(t-l_2\tau) \\
&+ \sum_{l_1=0}^{N_{l_3}-1} \sum_{l_2=0}^{N_{l_3}-1} \sum_{l_3=0}^{N_{l_3}-1} h_3(l_1, l_2, l_3)x(t-l_1\tau)x(t-l_2\tau)x(t-l_3\tau) + \dots
\end{aligned} \tag{3}$$

note $N_{\max} = \max(N_{l_1}, N_{l_2}, N_{l_3}, \dots, N_{l_i})$, ($i = 1, 2, 3, \dots$), when $n \geq N_{\max}$, the same to meet the input traffic flow signal $x_{l_i} = x(t-l_i\tau)$ is irrelevant to $y(t)$, then the formula (3) can be written as:

$$\begin{aligned}
x(t'+T) = h_0 &+ \sum_{l_1=0}^{N_{\max}-1} h_1(l_1)x(t-l_1\tau) + \sum_{l_1=0}^{N_{\max}-1} \sum_{l_2=0}^{N_{\max}-1} h_2(l_1, l_2)x(t-l_1\tau)x(t-l_2\tau) \\
&+ \sum_{l_1=0}^{N_{\max}-1} \sum_{l_2=0}^{N_{\max}-1} \sum_{l_3=0}^{N_{\max}-1} h_3(l_1, l_2, l_3)x(t-l_1\tau)x(t-l_2\tau)x(t-l_3\tau) + \dots
\end{aligned} \tag{4}$$

The power series expansion item of prediction results are in fact only related to Know from the above analysis of the traffic flow functional systems, the power series expansion item of prediction results are in fact only related to summation form all the products of the Input signal and the first power delay time signal. This means that the value of $N_{\max} = \max(N_{l_1}, N_{l_2}, N_{l_3}, \dots, N_{l_i})$, so $N_{\max} = \max(N_{l_1}, N_{l_2}, N_{l_3}, \dots, N_{l_i}) = m$.

3 Traffic Flow Time Series Volterra Neural Network Model

3.1 Representation of Nonlinear Systems Using Artificial Neural Network

Has proven that the BP neural network with one hidden layer can approximate any continuous bounded non-linear system, therefore, generally selected to contain a three-layer back propagation BP network with one hidden layer to approximate nonlinear systems. A single output three-layer back propagation neural network is shown in Figure 1. In the figure, the input vector $x_k^T = [x_{k,0}, x_{k,1}, \dots, x_{k,M}]$ at moment n can obtain by the delay of $x(k)$, where $x_{k,m} = x(k-m)$, the input of the l hidden unit($l = 1, 2, \dots, L$) is $Z_{l,k} = S_l(u_{l,k})$; $u_{l,k} = \sum_{m=0}^M w_{l,m} x_{k,m}$.

If the implicit function selected the sigmoid function, then $S_l(u_{l,k}) = \frac{1}{1 + \exp[-\lambda(u_{l,k} - \theta_l)]}$, where, θ_l is the threshold of the unit n. If the output unit is linear summation unit, the output at moment n is $y_k = \sum_{l=1}^L r_l Z_{l,k}$. The output of each hidden unit to expand into a Taylor series at the threshold θ_l : $Z_{l,k} = \varphi_l(u_{l,k}) = \sum_{i=0}^{\infty} d_i(\theta_l) u_{l,k}^i$, Where, $d_i(\theta_l)$ is the commencement of the

coefficient, the value associated with θ_l , and because of $u_{l,k} = \sum_{m=0}^M w_{l,m} x_{k,m}$, then the output of the neural network is

$$y_k = \sum_{l=1}^L r_l \sum_{i=0}^{\infty} d_i(\theta_l) \cdot \sum_{m_1=0}^M \cdots \sum_{m_i=0}^M w_{l,m_1} \cdots w_{l,m_i} x_{k,m_1} \cdots x_{k,m_i} \quad (5)$$

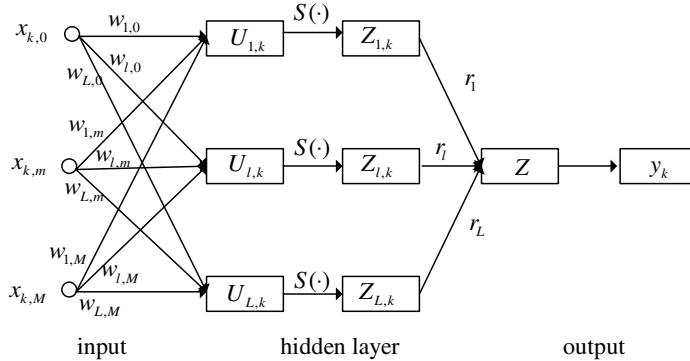


Fig. 1. Three layer neural networks in response to $M+1$ input and single output system

3.2 Traffic Flow Volterra Neural Network Model

Analysis and comparison of traffic flow Volterra series model in equation (4) and three-layer BP neural network in equation (5), then between them in the function, structure and method for solving are inherently close contact and similarity.

From a functional point of view, the traffic flow chaotic time series, Volterra series model and ANN model can be measured traffic flow chaotic time series, to simulate and predict the traffic flow process.

From a structural point of view, the traffic flow chaotic time series Volterra model and ANN model is also isomorphic.

From a method for solving point of view, Traffic flow Volterra model is based on orthogonal polynomials for the numerical approximation to find the approximate solution. Both have the same effect.

Through consistency of traffic flow chaotic time series Volterra model and ANN model, in this paper, the traffic flow chaotic time series Volterra neural network model (VNNTF) has been proposed in Figure 2. In the figure, $X(t) = (x(t), x(t+\tau), \dots, x(t+(m-1)\tau)^T$ ($t = 1, 2, \dots$) is the reconstructed phase space vector; $w_{i,j}$ ($i = 1, 2, \dots; j = 1, 2, \dots$); r_n is the traffic flow chaotic time series Volterra neural network weights parameters; g_s ($s = 1, 2, \dots, N$) is the activation function and $V_s(k)$ is the traffic flow of the convolution of the input signal:

$$V_N(t) = \sum_{i=0}^m w_{Ni} x(t+(i-1)\tau) \quad (6)$$

Thus, the traffic flow chaotic time series Volterra neural network expression is

$$\hat{y}(t) = f(X(t)) = f(\bar{x}(t)) = \sum_{s=1}^N r_s g_s(V_N(t)) = \sum_{s=1}^N r_s g_s \left(\sum_{i=0}^m w_{si} x(t + (i-1)\tau) \right) \quad (7)$$

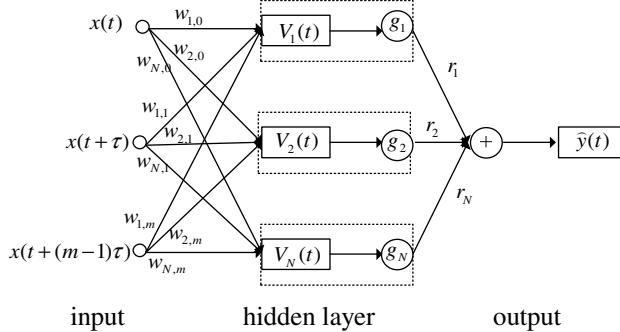


Fig. 2. The chaotic time series Volterra neural network traffic flow model (VNNTF)

4 VNNTF Model Rapid Learning Algorithm

The following traffic flow data used are from "Chongqing Road Traffic Management Data Sheet I and II" in 2006. There is the study of traffic volume time series of two-lane road 28 hours and 5 minutes every 5 minutes, and its sequence length $n = 337$. Then, the 4-9-1 structure of traffic flow Volterra neural network was obtained, specifically shown in Figure 2..

The steps of traffic flow chaotic time series Volterra Neural Network fast learning algorithm is showed and the specific steps are as follows:

Algorithm VNNTF model fast learning algorithm

Step1) The traffic flow chaotic time series Volterra Neural Network parameters

$w = (w_{s,i})_{N \times m}$ and r_s ($s = 1, 2, \dots, 9$, $i = 1, 2, \dots, 4$) are initialized.

Step2) Using phase space reconstruction theory to preprocess the traffic flow chaotic time series, and perform normalization for the reconstructed network input signal. Based on Takens theorem, the minimum embedding dimension $m = 4$, and the delay time $\tau = 3$. The reconstruction phase space vector number is $N - 1 - (m-1)\tau = 327$, which the top 250 vector are used as network input signals.

Step3) Using the initialized network and the preprocessed traffic flow time series, the first VNNTF neural network training begin with the function

$$\hat{y}(t) = \sum_{s=1}^N \sum_{i=1}^{+\infty} r_s a_{i,s} \left(\sum_{i=0}^m w_{si} x(t + (i-1)\tau) \right)^i .$$

Step4) Calculate error function, the function formula is $E(\theta) = \frac{1}{2} \sum_{t=1}^{250} (y(t) - \hat{y}(t))^2$. Set the maximum error is $E_{\max} = 0.035$, if $E < E_{\max}$,

the storage VNNTF neural network parameter use $w = (w_{s,i})_{N \times m}$ and r_s ($s = 1, 2, \dots, 9$, $i = 1, 2, \dots, 4$); and further $h_j(l_1, l_2, \dots, l_j)$ ($j = 1, 2, \dots, m$) can be calculated, otherwise, transferred to step5).

Step5) Calculate local gradient of the VNNTF. Specifically, according to the formula $\delta_j(t) = (y(t) - \bar{y}_j(t))g_s'(V_j(t))$ (j is the output layer) and the formula $\delta_j(t) = -\frac{\partial E(t)}{\partial y_j(t)}g_s'(V_j(t))$ (j is the hidden layer).

Step6) By introducing the momentum term, to adjust the learning weights of the traffic flow chaotic time series Volterra neural network. Parameter correction calculation formula is $\Delta w_{ji}^l(t+1) = -\eta \delta_j^{l+1}(t)x_i^l(t) + \alpha \Delta w_{ji}^l(t)$, where α is inertia factor; η is learning step; $\alpha \Delta w_{ji}(t+1)$ is the introduction of the momentum and $\delta_j(t)$ is calculated.

Step7) Calculating the modified weights in the traffic flow chaotic time series Volterra neural network in Step8) and transferred to step4), and train network again, then calculate the network output $\hat{y}(t)$ and the error E , repeated training until the relative error in traffic meet $E < E_{\max} = 0.035$.

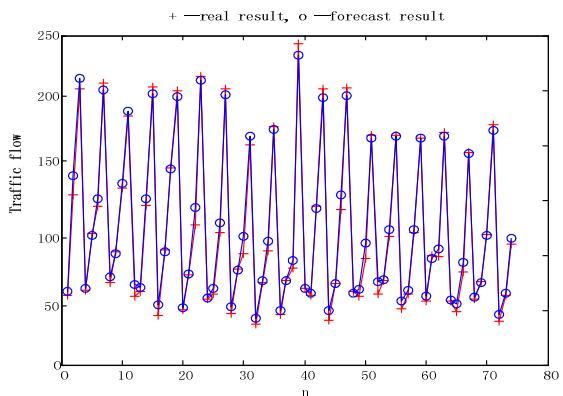
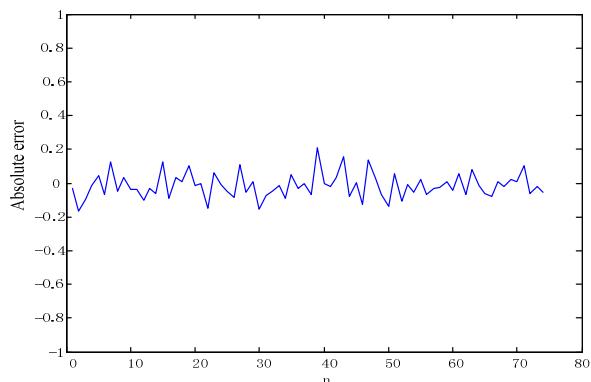
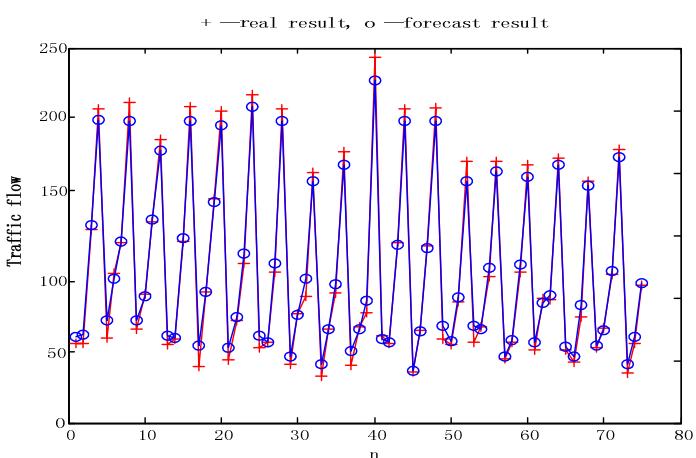
Step8) Output of every training storage of network parameters $w = (w_{s,i})_{N \times m}$ and r_s ($s = 1, 2, \dots, 9$, $i = 1, 2, \dots, 4$). The activation function $g_s(V_s(t))$ is expanded into a Taylor series at the threshold θ_s and $d_i(\theta_s)$ is obtained. If the activation function is a polynomial, then $d_i(\theta_s) = a_{i,s}$ ($s = 1, 2, \dots, 9$, $i = 1, 2, \dots, 4$).

Step9) According to the formula $h_j(l_1, l_2, \dots, l_j) = \sum_{s=1}^N r_s d_i(\theta_s) w_{s,l_1} w_{s,l_2} \dots w_{s,l_j}$, the kernel function ($s = 1, 2, \dots, 9$, $i = 1, 2, \dots, 4$) of the output system is calculated.

5 Experimental Results and Analysis

Multi-step prediction is a major aspect to reflect the performance of predictive model. Traffic flow time series Multi-step prediction is as follows: If the sample size is N , in the new data point can not be used or only the sample points N , It can be predicted beyond $N+1$ values, can also predict the $N+2$ values, $N+3$ values ... $N+T$ values ($T > 0$).That is, for the known sample set can be extrapolated to predict T step. The following multi-step prediction of the traffic flow VNNTF network, and the results compare with the multi-step prediction of the BP neural network and filter Voltrra, further, analyzing the causes of the different predictions.

Were calculated the error root mean square in Figure 5, 7 and 9, and these results are compared with the BP network and the wavelet neural network based on the non-chaotic algorithm, and the compare results are shown in Table 1. From Table 1, with the increasing number of prediction steps, in which the same prediction step, the root mean square of the wavelet neural network based on chaotic algorithm is significantly less than the BP neural network and the wavelet neural network based on non-chaotic algorithm.

**Fig. 4.** The 2-step forecast result and real result**Fig. 5.** The 2-step forecast error curve**Fig. 6.** The 3-step forecast result and real result

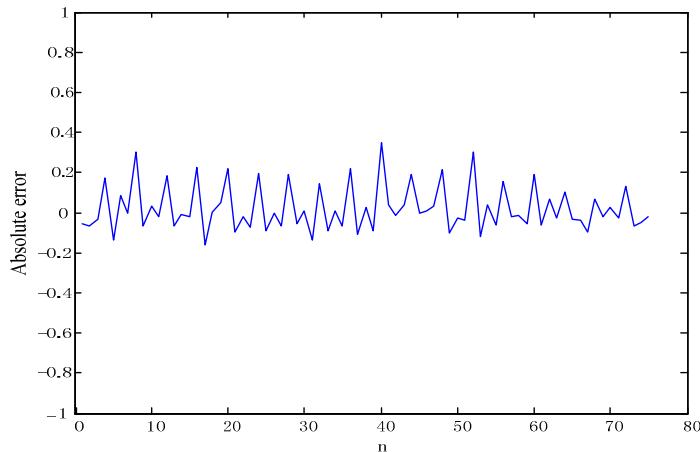


Fig. 7. The 3-step forecast error curve

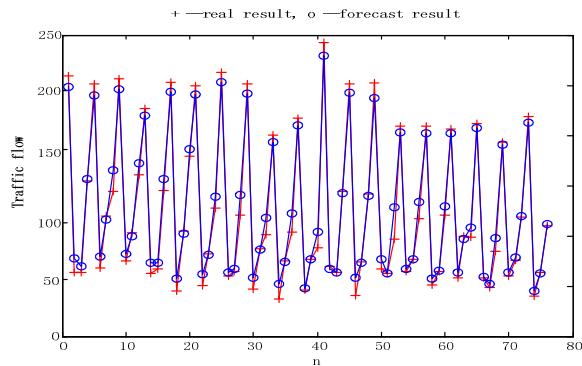


Fig. 8. The 4-step forecast result and real result

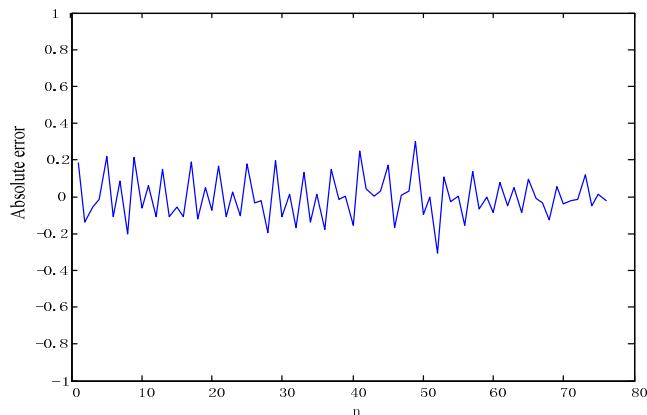


Fig. 9. The 4-step forecast error curve

Figure 4, 6 and 8, respectively, which corresponds to 2-step, 3-step and 4-step predicted and actual comparison curves of VNNTF network based VNNTF network rapid learning algorithm; and “+” shows the true value, “o” shows the forecasted value. Figure 3 , 7, and 9 respectively, which corresponds to 2-step, 3-step and 4-step absolute error curve .Figure 4 to Figure 9 shows the general trend is to predict the longer the step, the prediction performance has become getting worse.

Table 1. Normalization of RMSE comparison

prediction step	BP network	Volterra filter	VNNTF network
1 step	0.7014	0.3567	0.1368
2 step	0.8074	0.3941	0.1507
3 step	0.8653	0.4225	0.2322
4 step	0.9799	0.4782	0.2417

Analysis of multi-step prediction results to VNNTF network, the 2-step, 3-step and 4-step predictable performance overall is better than the BP neural network prediction and the Volterra filter prediction; this is because the network VNNTF combines the Volterra series and ANN network advantages, to overcome the difficulties of solving the Volterra kernel function and the blindness of ANN network modeling.

6 Conclusions

In the paper traffic flow chaotic time series VNNTF model was designed. A traffic flow VNNTF fast learning algorithm based on chaos theory was proposed. The method of model selection and algorithm design, are considered the chaotic characteristics of traffic flow time series, which is a theoretical value. Simulation results show that the method can reduce network training time and improve the forecast accuracy, and show better predictive effectiveness and reliability.

Acknowledgments. This research is financially supported by the National Natural Science Funds of China for Distinguished Young Scholar under Grant(50925727), and the Fundamental Research Funds for the Central Universities, Hefei University of Technology for Professor He Yigang , the National Natural Science Foundation of China(NSFC) for Professor Xue-ping Dong (No. 60974022) and the Universities Natural Science Foundation of Anhui Province (No. KJ2012A219) for Professor Yin Lisheng.

References

1. Asyali, M., Alc, M.: Obtaining Volterra Kernels from Neural Networks. World Congress on Medical Physics and Biomedical Engineering 2, 11–15 (2006)
2. Ghasemi, M., Tavassoli Kajani, M., Babolian, E.: Numerical solutions of the nonlinear Volterra–Fredholm integral equations by using homotopy perturbation method. Applied Mathematics and Computation 188(1), 446–449 (2007)

3. Liu, B., Zhang, Y., Chen, L.: Dynamic complexities in a lotka–volterra predator–prey model concerning impulsive control strategy. *International Journal of Biomathematics* 1(1), 179–1964 (2008)
4. Yakubov, A.Y.: On nonlinear Volterra equations of convolution type. *Differential Equations* 45(9), 1326–1336 (2009)
5. Kobayakawa, S., Yokoi, H.: Evaluation of Prediction Capability of Non-recursion Type 2nd-order Volterra Neuron Network for Electrocardiogram. In: Köppen, M., Kasabov, N., Coghill, G. (eds.) *ICONIP 2008, Part II. LNCS*, vol. 5507, pp. 679–686. Springer, Heidelberg (2009)
6. Kang, L., Wang, C., Jiang, T.: Hydrologic model of Volterra neural network and its application. *Journal of Hydroelectric Engineering* 25(5), 22–26 (2006)
7. Yuan, H., Chen, G.: Fault Diagnosis in Nonlinear Circuit Based on Volterra Series and Recurrent Neural Network. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) *ICONIP 2006. LNCS*, vol. 4234, pp. 518–525. Springer, Heidelberg (2006)
8. Si, W., Duan, Z.-M., Wang, H.-T.: Novel Method Based on Projection of Vectors in Linear Space to Identify Volterra Kernels of Arbitrary Orders. *Application Research of Computers* 25(11), 3340–3342 (2008)
9. Yakubov, A.Y.: On nonlinear Volterra equations of convolution type. *Differential Equations* 45(9), 1326–1336 (2009)
10. Murakami, S., Huu, P., Ngoc, A.: On stability and robust stability of positive linear Volterra equations in Banach lattices. *Central European Journal of Mathematics* 8(5), 966–984 (2010)
11. Bibik, Y.V.: The second Hamiltonian structure for a special case of the Lotka–Volterra equations. *Computational Mathematics and Mathematical Physics* 47(8), 1285–1294 (2007)
12. Yin, L.-S., Huang, X.-Y., Yang, Z.-Y., Xiang, C.-C.: Prediction for Chaotic Time Series Based on Discrete Volterra Neural Networks. In: Wang, J., Yi, Z., Žurada, J.M., Lu, B.-L., Yin, H. (eds.) *ISNN 2006. LNCS*, vol. 3972, pp. 759–764. Springer, Heidelberg (2006)

Improved Calculation Scheme of Structure Matrix of Boolean Network Using Semi-tensor Product

Jinyu Zhan*, Shan Lu, and Guowu Yang

College of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 610054, China
`{zhanjy,guowu}@uestc.edu.cn, 617999242@qq.com`

Abstract. Semi-tensor product is an efficient tool for analyzing the characteristics of a Boolean network. Using it, the number of fixed points and numbers of all possible circles of different lengths are determined by the structure matrix of a Boolean network. But the conventional method to obtain the structure matrix is very complex. In this paper, a novel method is proposed to get the structure matrix. Unlike existing methods, our approach gets the structure matrix of a Boolean network not through the complex matrix operations but through the truth table which reflects the state transformation of the Boolean network. Comparing our solutions with the conventional method shows the advantage of our new approach through an example of a Boolean Network.

Keywords: Semi-tensor Product, Boolean Network, Structure Matrix, Truth Table.

1 Introduction

The Boolean network, introduced firstly by Kauffman [1], and then developed by [2][3][4][5][6][7][8] and many others, becomes a powerful tool in describing, analyzing, and simulating the cell network. It was shown that the Boolean network plays an important role in modeling cell regulation, because they can represent important features of living organisms [9][10]. It has received the most attention, not only from the biology community, but also physics, system science, etc.

The structure of a Boolean network is described in terms of its cycles and the transient states that lead to them. Several useful Boolean networks have been analyzed and their circles have been revealed[11][12]. It was pointed in [13] that finding fixed points and circles of a Boolean network is an NP hard problem. Semi-tensor product of matrix(STP), presented by Cheng[14]. Using STP, a Boolean network equation can be expressed as a conventional discrete time linear system which contains complete information of the dynamics of a

* This work was supported by the Fundamental Research Funds for the Central Universities of China under Grant No. ZYGX2009J062 and the National Natural Science Foundation of China under Grant No. 60973016.

Boolean network. Analyzing the structure matrix of a Boolean network, precise formulas are obtained to determine the number of fixed points and numbers of all possible circles of different lengths.

But the conventional method to calculate the structure matrix of a Boolean network, presented in [15][18][19][21], is very complex. In this paper, a novel method is proposed to get the structure matrix. Unlike existing methods, our approach gets the structure matrix of a Boolean network not through the complex matrix operations but through the truth table which reflects the state transformation of the Boolean network. Compared with the conventional method, our method can greatly reduce the calculation complexity. An example of a Boolean network illustrates the effectiveness of our method.

The rest of the paper is organized as follows. Section 2 gives a brief introduction to semi-tensor product of matrices, matrix expression of logic and dynamics of Boolean network. The conventional method to calculate the structure Matrix of a Boolean Network is given in Section 3 and our method is proposed in Section 4. Finally, some conclusions are drawn in Section 5.

2 Expression of Boolean Network in Semi-tensor Product

2.1 Semi-tensor Product

This section is a brief introduction to semi-tensor product(STP) of matrices. STP of matrices is a generalization of conventional matrix product, which extends the conventional matrix product to any two matrices. It plays a fundamental rule in the following discussion. We restrict it to some concepts and basic properties used in this paper. In addition, only left semi-tensor product for multiplying dimension case is involved in the paper. We refer to [14][15][16][17] for right semi-tensor product, arbitrary dimensional case and much more details. Throughout this paper "semi-tensor product" means the left semi-tensor product for multiplying dimensional case.

Definition 1. 1. Let X be a row vector of dimension np , and Y be a column vector with dimension p . Then we split X into p equal-size blocks as X^1, X^2, \dots, X^p , which are $1 \times n$ rows. Define the STP, denoted by \bowtie , as:

$$\begin{cases} X \bowtie Y = \sum_{i=1}^p X^i y_i \in \mathbb{R}^n, \\ Y^T \bowtie X^T = \sum_{i=1}^p y_i (X^i)^T \in \mathbb{R}^n. \end{cases} \quad (1)$$

2. Let $A \in M_{m \times n}$ and $B \in M_{p \times q}$. If either n is a factor of p , say $nt = p$ and denote it as $A \prec_t B$, or p is a factor of n , say $n = pt$ and denote it as $A \succ_t B$, then we define the STP of A and B , denoted by $C = A \times B$, as the following: C consists of $m \times q$ blocks as $C = (C^{ij})$ and each block is $C^{ij} = A^i \times B_j$, where $i = 1, \dots, m$, $j = 1, \dots, q$, A^i is the i -th row of A and B_j is the j -th column of B .

we use some simple numerical examples to describe it.

Example 1. Let $X = [1 \ 2 \ 3 \ -1]$ and $Y = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Then

$$X \times Y = [1 \ 2] \cdot 1 + [3 \ -1] \cdot 2 = [7 \ 0]$$

2.2 Matrix Expression of Logic

In this section, the matrix expression of logic will be given. In a logical domain, we usually set "true" as "1" and "false" as "0". Then a logical variable is defined as $x \in D = 0, 1$. There are several fundamental binary functions such as \neg , \wedge , \vee , \leftrightarrow , \rightarrow , $\bar{\vee}$, \uparrow and \downarrow . Their truth table is as Table 1.

Table 1. Truth table of \neg , \wedge , \vee , \leftrightarrow , \rightarrow , $\bar{\vee}$, \uparrow and \downarrow

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \leftrightarrow q$	$p \rightarrow q$	$p \bar{\vee} q$	$p \uparrow q$	$p \downarrow q$
0	0	1	0	0	1	1	0	1	1
0	1	1	0	1	0	1	1	1	0
1	0	0	0	1	0	0	1	1	0
1	1	0	1	1	1	1	0	0	0

To use matrix expression each element can be identified in D with a vector as $1 \sim \delta_2^1$ and $0 \sim \delta_2^2$, where $\delta_n^i = Col(I_n)$. Therefore, that a n-ary logical operator (or function) is a mapping: $f : D^n \rightarrow D$ can be formed as $f : \Delta_{2^n} \rightarrow \Delta$.

Theorem 1. Let $f(x_1, \dots, x_n)$ be a logical function in vector form as $f : \Delta_{2^n} \rightarrow \Delta$. Then there exists an unique $M_f \in L_{2 \times 2^n}$, called the structure matrix of f , such that

$$f(x_1, \dots, x_n) = M_f \times x, \quad \text{where } x = \times_{i=1}^n x_i. \quad (2)$$

Therefore, the structure matrix of \neg , \wedge , \vee , \leftrightarrow , \rightarrow , $\bar{\vee}$, \uparrow and \downarrow are as follows: $M_{\neg} = \delta_2[2 \ 1]$, $M_1 = \delta_2[1 \ 2 \ 2 \ 2]$, $M_{\vee} = \delta_2[1 \ 1 \ 1 \ 2]$, $M_{\leftrightarrow} = \delta_2[1 \ 2 \ 2 \ 1]$, $M_{\rightarrow} = \delta_2[1 \ 2 \ 2 \ 1]$, $M_{\bar{\vee}} = \delta_2[2 \ 1 \ 1 \ 2]$, $M_{\uparrow} = \delta_2[2 \ 1 \ 1 \ 1]$, $M_{\downarrow} = \delta_2[2 \ 2 \ 2 \ 1]$.

Theorem 2. Let $F(x_1, \dots, x_n) : D^n \rightarrow D^k$ be a logical mapping: $F : \Delta_{2^n} \rightarrow \Delta_{2^k}$. Then there exists an unique $M_F \in L_{2^k \times 2^n}$ called the structure matrix of F , such that

$$F(x_1, \dots, x_n) = M_F \times x \quad (3)$$

2.3 Dynamics of Boolean Networks

The Boolean networks play an important role in modeling cell regulation, because they can represent important features of living organisms. The dynamics of the Boolean networks will be given in this section.

Definition 2 [15][18][20]. A Boolean network is a set of nodes A_1, A_2, \dots, A_n , which interact with each other in a synchronous manner. At each given time $t = 0, 1, 2, \dots$, a node has only one of two different values: 1 or 0. Thus the network can be described by a set of equations:

$$\begin{cases} A_1(t+1) = f_1(A_1(t), A_2, \dots, A_n(t)) \\ A_2(t+1) = f_2(A_2(t), A_2, \dots, A_n(t)) \\ \vdots \\ A_n(t+1) = f_n(A_1(t), A_2, \dots, A_n(t)) \end{cases} \quad (4)$$

Where f_i , ($i = 1, 2, \dots, n$), are n-ary logic functions.

Note that in Boolean networks each function f_i has only constant, linear, or product terms[12].

3 Conventional Calculation of the Structure Matrix

Using Theorem 1 and 2, the dynamics of Boolean networks can be expressed as $A(t+1) = LA(t)$, where $A(t+1) = \times_{i=1}^l A_i(t+1)$, $A(t) = \times_{i=1}^l A_i(t)$, L is the structure matrix of F , $L \in \mathbb{L}_{2^l \times 2^l}$.

By means of the STP, the dynamics of Boolean networks can be converted into the equivalent algebraic forms. Through the analysis of the structure matrix L , we can get the characteristics of the Boolean networks such as: (1)fixed points; (2)circles of different lengths; (3)transient period; (4) basin of each attractor[15][18]. Therefore, how to get the structure matrix L easily is very important. The conventional method to get the structure matrix L is as follows.

Firstly, a simple example is given to show the structure of a Boolean network.

Example 2. Consider a Boolean network which dynamics is described as

$$\begin{cases} A_1(t+1) = \overline{A_2(t) \cdot A_3(t)} \cdot A_1(t) \\ A_2(t+1) = A_1(t) \cdot \overline{A_2(t)} + \overline{A_1(t)} \cdot \overline{A_3(t)} \cdot A_2(t) \\ A_3(t+1) = A_1(t) \cdot A_2(t) \cdot \overline{A_3(t)} + \overline{A_2(t)} \cdot A_3(t) \end{cases} \quad (5)$$

In algebraic form(the notation \times is omitted), we can have

$$\begin{cases} A_1(t+1) = M_{\downarrow}(M_{\wedge} A_2(t) A_3(t)) A_1(t) \\ A_2(t+1) = M_{\vee}((M_{\wedge} A_1(t)(M_{\neg} A_2(t))(M_{\wedge} (M_{\downarrow} A_1(t) A_3(t)) A_2(t))) \\ A_3(t+1) = M_{\vee}((M_{\wedge} (M_{\wedge} A_1(t) A_2(t))(M_{\neg} A_3(t)))(M_{\wedge} (M_{\neg} A_2(t)) A_3(t))) \end{cases} \quad (6)$$

Using some theorems of the paper[15][18] and $A(t+1) = LA(t)$, the structure matrix L is as follows:

$$\begin{aligned}
L = & M_{\vee} M_{\wedge} M_{\wedge} (I_2 \otimes (I_2 \otimes M_{\neg}(I_2 \otimes M_{\wedge} M_{\neg}(I_2 \otimes (I_2 \otimes M_{\vee} M_{\wedge}(I_2 \otimes M_{\neg} \\
& (I_2 \otimes M_{\wedge} M_{\downarrow}(I_2 \otimes (I_2 \otimes (I_2 \otimes M_{\downarrow} M_{\wedge})))))))))) W_{[2]}(I_2 \otimes W_{[2]})(I_4 \otimes \\
& W_{[2]})(I_8 \otimes W_{[2]})(I_{32} \otimes W_{[2]})(I_{128} \otimes W_{[2]})(I_{256} \otimes W_{[2]})(I_{512} \otimes W_{[2]} \\
& (I_{1024} \otimes W_{[2]})W_{[2]}(I_4 \otimes W_{[2]})(I_{16} \otimes W_{[2]})(I_{64} \otimes W_{[2]})(I_{128} \otimes W_{[2]}) \\
& (I_{256} \otimes W_{[2]})(I_{512} \otimes W_{[2]})(I_2 \otimes W_{[2]})(I_{32} \otimes W_{[2]})(I_{64} \otimes W_{[2]})(I_{128} \otimes \\
& W_{[2]})(I_{256} \otimes W_{[2]})(I_{16} \otimes W_{[2]})(I_{128} \otimes W_{[2]})(I_8 \otimes W_{[2]})(I_{64} \otimes W_{[2]}) \\
& (I_4 \otimes W_{[2]})(I_{32} \otimes W_{[2]})(I_{16} \otimes W_{[2]})(I_8 \otimes W_{[2]})M_r M_r M_r (I_2 \otimes (M_r M_r \\
& M_r M_r (I_2 \otimes M_r M_r M_r)))
\end{aligned} \tag{7}$$

The process is very complex and the biggest order of the matrices in the equation is more than 1024.

4 New Method for the Calculation of Structure Matrix

The conventional method to calculate the structure matrix L is very complex. A new method will be proposed in this section.

Definition 3. Form a square matrix by all the present-state vectors $A(t) = \times_{i=1}^l A_i(t)$, the matrix is called present-state matrix, denoted by $Q(t)$. There is another matrix correspond to $Q(t)$, called next-state matrix, denoted by $Q(t+1)$.

As $A(t+1) = L A(t)$, we can derive $Q(t+1) = L Q(t)$. It is easy to know that $Q(t) \in \mathbb{L}_{2^l \times 2^l}$, and $Q(t)$ is an invertible matrix. Then the structure matrix $L = Q(t+1)[Q(t)]^{-1}$. Further simplify the calculation, $Q(t)$ can be arrayed to 2^l -order identity matrix. Therefore, $L = Q(t+1)$.

For the example 2, we have the truth table as Table 2.

Table 2. Truth table of the example 2

$A_3(t)$	$A_2(t)$	$A_1(t)$	$A_3(t+1)$	$A_2(t+1)$	$A_1(t+1)$
0	0	0	0	0	1
0	0	1	0	1	0
0	1	0	0	1	1
0	1	1	1	0	0
1	0	0	1	0	1
1	0	1	1	1	0
1	1	0	0	0	0
1	1	1	0	0	0

The state vectors' table is as Table 3.

Table 3. The state vectors' table

$A_3(t)$	$A_2(t)$	$A_1(t)$	$A(t)$
0	0	0	[0 0 0 0 0 0 0 1] ^T
0	0	1	[0 0 0 0 0 0 1 0] ^T
0	1	0	[0 0 0 0 0 1 0 0] ^T
0	1	1	[0 0 0 0 1 0 0 0] ^T
1	0	0	[0 0 0 1 0 0 0 0] ^T
1	0	1	[0 0 1 0 0 0 0 0] ^T
1	1	0	[0 1 0 0 0 0 0 0] ^T
1	1	1	[1 0 0 0 0 0 0 0] ^T

Then, we can get the present-state matrix and next-state matrix as follows:

$$Q(t) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad Q(t+1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Therefore, the structure matrix $L = Q(t+1)$.

By our method, it is easy to get the structure matrix through the truth table which reflects the transformation of the states. Our method to get the structure matrix is simpler than the conventional method.

5 Conclusion

The structure matrix is very important tool to analyzing fixed points and circles of a Boolean network. In the conventional method, the structure matrix is calculated through complex matrix operations. In this paper, a new method is proposed to get the structure matrix. Unlike existing methods, our approach gets the structure matrix of a Boolean network through the truth table which reflects the state transformation of the Boolean network. Compared with the conventional method, our method can greatly reduce the calculation complexity. An example of a Boolean network illustrates the advantage of our method.

Finding fixed points and circles of a Boolean network is an NP hard problem. A method using semi-tensor product is proposed by cheng[15][18][21]. But the calculation is also complex. Therefore, we will deal with the problems in the future work.

References

1. Kauffman, S.A.: Metabolic stability and epigenesist in randomly constructed genetic nets. *Journal of Theoretical Biology* 22, 437–467 (1969)
2. Akutsu, T., Miyano, S., Kuhara, S.: Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16, 727–773 (2000)
3. Albert, R., Barabasi, A.L.: Dynamics of complex systems: scaling laws or the period of Boolean networks. *Phys. Rev. Lett.* 84, 5660–5663 (2000)
4. Shumlevich, I., Dougherty, R., Kim, S., Zhang, W.: Probabilistic Boolean network: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18(2), 261–274 (2002)
5. Harris, S.E., Sawhill, B.K., Wuensche, A., Kauffman, S.: A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity* 7, 23–40 (2002)
6. Aldana, M.: Boolean dynamics of networks with scale-free topology. *Physica D* 185, 45–66 (2003)
7. Samuelsson, B., Troein, C.: Superpolynomial growth in the number of attractors in Kauffman networks. *Phys. Rev. Lett.* 90, 90098701 (2003)
8. Drossel, B., Mihaljev, T., Greil, F.: Number and length of attractors in a critical Kauffman model with connectivity one. *Phys. Rev. Lett.* 94, 088701 (2005)
9. Albert, R., Othmer, H.G.: The topology and signature of the regulatory interactions predict the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *Journal of Theoretical Biology* 223(1), 1–18 (2003)
10. Huang, S.: Regulation of cellular states in mammalian cells from a genomewide view. In: Collado-Vides, J., Hofestadt, R. (eds.) *Gene Regulation and Metabolism*, pp. 181–220. MIT Press, Cambridge (2002)
11. Heidel, J., Maloney, J., Farrow, J., Rogers, J.: Finding cycles in synchronous Boolean networks with applications to biochemical systems. *Int. J. Bifurcat. Chaos* 13(3), 535–552 (2003)
12. Farrow, C., Heidel, J., Maloney, H., Rogers, J.: Scalar equations for synchronous Boolean networks with biological applications. *IEEE Trans. Neural Networks* 15(2), 348–354 (2004)
13. Zhao, Q.: A remark on Scalar Equations for synchronous Boolean Networks with biologic applications by C.Farrow, J.Heidel, J.Maloney, and J.Rogers. *IEEE Trans. Neural Networks* 16(6), 1715–1716 (2005)
14. Cheng, D., Qi, H.: Semi-tensor Product of Matrices-Theory and Application, 2nd edn. Science Press, Beijing (2011)
15. Cheng, D., Qi, H., Li, Z.: Analysis and Control of Boolean Networks: A Semi-tensor Product Approach. Springer Press, London (2011)
16. Cheng, D.: Matrix and Polynomial Approach to Dynamic Control Systems. Science Press, Beijing (2002)
17. Cheng, D.: Semi-tensor product of matrices and its applications - A survey. In: ICCM 2007, vol. 3, pp. 641–668 (2007)
18. Cheng, D., Qi, H., Zhao, Y.: Analysis and control of Boolean networks: a semi-tensor product approach. *Acta Automatica Sinica* 37(5), 529–539 (2011)
19. Cheng, D., Qi, H., Li, Z.: Model construction of Boolean network via observed data. *IEEE Transactions on Neural Networks* 22(4), 525–536 (2011)
20. Cheng, D., Qi, H.: A linear representation of dynamics of boolean networks. *IEEE Transactions on Automatic Control* 55(10), 2251–2258 (2011)
21. Cheng, D., Qi, H.: State-space analysis of Boolean networks. *IEEE Transaction on Neural Networks* 21(4), 584–594 (2010)

Adaptive RBF Neural Network Filtering Predictive Model Based on Chaotic Algorithm

Lisheng Yin, Yigang He, Xueping Dong, and Zhaoquan Lu

School of Electrical and Automation Engineering, Hefei University of Technology,
230009, Hefei, China

yls20000@163.com, hyghnu@yahoo.com.cn, {hfdxp, luzhquan}@126.com

Abstract. In this paper, based on the RBF neural networks and the deterministic and nonlinear characterization of chaotic time series, the adaptive RBF neural network filtering predictive model based on chaotic algorithm is proposed to make prediction of chaotic time series. The predictive model of chaotic time series is established with the adaptive RBF neural networks and the steps of the chaotic learning algorithm are expressed. The network system can enhance the stabilization and associative memory of chaotic dynamics and generalization ability of predictive model even by imperfect and variation inputs during the learning and prediction process by selecting the suitable nonlinear feedback term. The model is tested for the chaotic time series which venerated with Lorentz system by on-line method. The experimental and simulating results indicated that the adaptive RBF neural network filtering predictive model has a good adaptive prediction performance and can be successfully used to predict chaotic time series.

Keywords: RBF neural network, chaotic algorithm, chaotic time series, prediction.

1 Introduction

Prediction for chaotic time series is to approximate the unknown nonlinear functional mapping of a chaotic signal. The laws underlying the chaotic time series can be expressed as a deterministic dynamical system. Farmer and Sidorowich suggest reconstructing the dynamics in phase space by choosing a suitable embedding dimension and time delay[1]. Takens' theorem ensures that the method is reliable, based on the fact that the interaction between the variables is such that every component contains information on the complex dynamics of the system[2].

In recent years, particular interest has been put into predicting chaotic time series using neural networks because of their universal approximation capabilities. Most applications in this field are based on feed-forward neural networks, such as the Back Propagation (BP) network [3-6], Radial Basis Function (RBF) network[7-8], Recurrent neural networks (RNNs)[9-11], FIR neural networks [12] and so on. It is widely used tool for the prediction of time series[13-15].

In this paper, the chaotic algorithm is proposed to a RBF neural network filtering predictive model and the model is proposed to make prediction of chaotic time series. The model is tested for the chaotic time series which venerated with Lorentz system by on-line method. The experimental and simulation results indicated that the adaptive filtering has a good self-suitable prediction performance and can be successfully used to predict chaotic time series.

2 Establishment of Adaptive RBF Neural Network Filtering Predictive Model Based on Chaotic Algorithm

2.1 Model of Chaotic Time Series Prediction

The chaotic time series prediction is based on the Takens' delay-coordinate phase reconstruct theory. If the time series of one of the variables is available, based on the fact that the interaction between the variables is such that every component contains information on the complex dynamics of the system, a smooth function can be found to model the portraits of time series. If the chaotic time series are $\{x(t)\}$, then the reconstruct state vector is $x(t) = (x(t), x(t+\tau), \dots, x(t+(m-1)\tau))$, Where m ($m = 2, 3, \dots$) is called the embedding dimension ($m = 2d+1$, d is called the freedom of dynamics of the system), and τ is the delay time. The predictive reconstruct of chaotic series is a inverse problem to the dynamics of the system essentially. There exists a smooth function defined on the reconstructed manifold in R^m to interpret the dynamics $x(t+T) = F(x(t))$, where T ($T > 0$) is forward predictive step length, and $F(\cdot)$ is the reconstructed predictive model.

2.2 Realized Architecture of Adaptive RBF Neural Network Filtering Predictive Model

After reconstructing the phase space, the RBF neural networks adopt three layers networks of Fig. 1. Where the input layer has m nerve cells, the first layer feed to the second layer directly and it do not need the power processing. r_i ($i = 1, 2, \dots, L$) is the reference vector and ϖ_k ($i = 1, 2, \dots, L$) is the adjustable parameters in the adaptive RBF neural network filtering. Thus, the adaptive RBF neural network filtering is more flexible in studying the nonlinear functions. The differentiation between the networks and the traditional neural networks is that the activation function is a RBF function but not the Sigmoid function. The activation function usually choose the Gauss function, the spline function $f(d_i(k))$, where $d_i(k) = \|x(k) - r_i(k)\|$. In the adaptive RBF neural network filtering, $y(k)$ is expressed as $\hat{y}(k) = f_2(\sum_{i=0}^{L-1} \varpi_i(k) f(d_i(k)))$, $i = 0, 2, \dots, L-1$, where $f_2(\cdot)$ is the activation function of output signal.

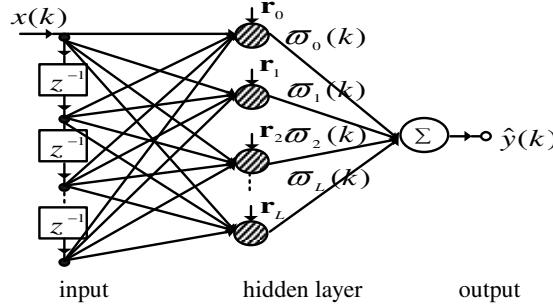


Fig. 1. Structure of adaptive RBF neural network filtering

Generally, the learning of the RBF neural network filtering has three steps. If the gradient method and the Gauss activation function are adopted, the regulate formulas of RBF are shown as:

$$\begin{cases} \varpi_i(k+1) = \varpi_i(k) + 2\mu_\varpi e(k) f(d_i(k)) \\ \sigma_i(k+1) = \sigma_i(k) + 2\mu_\sigma e(k) f(d_i(k)) \varpi_i(k) \frac{d_i^2(k)}{\sigma_i^3(k)} \\ \mathbf{r}_i(k+1) = \mathbf{r}_i(k) + 2\mu_r e(k) f(d_i(k)) \varpi_i(k) \frac{\mathbf{x}(k) - \mathbf{r}_i(k)}{\sigma_i^2(k)} \end{cases} \quad (1)$$

$$i = 0, 1, 2, \dots, L-1.$$

The RBF neural network system can enhance the stabilization and associative memory of chaotic dynamics and generalization ability of predictive model even by imperfect and variation inputs by selecting the suitable nonlinear feedback term. The dynamics of network become chaotic one in the weight space. Thus, the regulate formula $\varpi(k)$ is shown as

$$\varpi_i(k+1) = \varpi_i(k) + 2\mu_\varpi e(k) f(d_i(k)) + g(\varpi_i(k) - \varpi_i(k-1)) \quad (2)$$

where $g(x) = \tanh(ax) \exp(-bx^2)$, $x = \varpi_i(k) - \varpi_i(k-1)$.

That the feedback function $g(x)$ is chose is because that $g(x)$ can get the difference feedback function corresponding to the dissimilar parameter, such as the staircase function, δ function and so on. If the feedback function is seen as the motion-promoting force, the different feedback parameters a and b corresponding to the amplitude and width of the motion-promoting force. The paper [16] was detailed to discuss the influences by selecting the suitable learning and predictive process.

3 Adaptive RBF Neural Network Rapid Learning Algorithm

On the establishment of chaotic time series RBF, Network input the number of neurons, hidden layers and the number of neurons in the hidden layer are to be considered. The following chaotic time series used are from Lorenz chaotic sampling time series. The number of hidden layer wavelet neural taken as 9 by Kolmogorov Theorem, the number of input layer neurons equal to the minimum embedding dimension, the number of output layer is 1, so that the 4-9-1 structure of Lorenz chaotic sampling time series RBF was obtained, specifically shown in Figure 1.

Algorithm. The steps of the chaotic time series learning and prediction of the adaptive RBF neural network filtering predictive model are showed:

Step1) The dimension m of chaotic time series is calculated by the way of G- P algorithms, and the delay time τ is calculated by the self-correlation method. For the overall description of the dynamics characteristic of the original system by the Takens' delay-coordinate phase reconstruct theory, a chaotic series demand $m \geq 2d + 1$ variances at least, so the number of the input nerve cells of the adaptive RBF neural network filtering is $M = m$; The reconstruction phase space vector number is 200, Then, the 200 phase space vectors to make a simple normalized.

Step2) The adaptive filtering is initialized and the weights are vested the initial values. RBF neural network vector weighting parameters w is initialized, where the weight vector w in each component take random function between 0 and 1; and the learning rate η is initialized at the same time, where $\eta = 0.0002 \cdot \beta$ and γ are the learning rate adjustment factors, where $\beta = 0.75, \gamma = 1.05$.

Step3) Using the above the initialization network and the pretreatment traffic flow time series, the first training network is carried out.

Step4) If the error is in the scope of the permission, the error is calculated and it turns into Step4), otherwise it continues; the error function formula:

$$E(\theta) = \frac{1}{2} \sum_{t=1}^{250} (y(t) - \hat{y}(t))^2 \quad (3)$$

Set the maximum error is $E_{\max} = 0.035$, if $E < E_{\max}$, the storage RBF neural network parameter use w ; otherwise, then a second training network will be required.

Step5) If A previous training error is recorded as E_{n-1} , the current error is recorded as E_n , then Calculate the ratio of E_n to E_{n-1} , Setting constants $k = 1.04$, if $E_n / E_{n-1} > k = 1.04$, then substitute $\beta\eta$ for η to reduce learning rate; otherwise, replace η with $\gamma\eta$ to increase learning rate.

Step6) In the adaptive RBF neural network filtering for the chaotic time series prediction in Fig.1, $x(k) = x(t) t = 1, 2, \dots, N$ is the input, $\hat{y}(k) = \hat{x}(t)$ is the output.

Introduce nonlinear feedback into the weighting formal to adopt Chaos Mechanisms, due to the nonlinear feedback is vector form of weighting variables. In order to facilitate understanding, respectively, gives the vector w and its weighting formal, as follows. Parameter correction calculation formula is $\Delta w_{ji}^l(t+1) = -\eta \delta_j^{l+1}(t) x_i^l(t) + \alpha \Delta w_{ji}^l(t)$,

where α is inertia factor; η is learning step; $\alpha\Delta w_{j,i}(t+1)$ is the introduction of the momentum

Step7) Using the new learning rate in Step5) and RBF network parameters with nonlinear feedback in Step6) to calculate the new value ,and train network again , then get the error and enter into Step4, repeated training until the relative error in traffic meet $E < E_{\max}$.

Step8) Output of each stored network parameters and training error curve.

4 Example Analysis and Conclusions

4.1 Model and Data

Considering Lorenz chaotic system

Lorenz map:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = rx - y - xz \\ \dot{z} = -bz + xy \end{cases} \quad (4)$$

Where $\sigma = 10$, $r = 28$, $b = 8/3$.The initial value is $x(0) = 0$, $y(0) = 5$, $z(0) = -5$; and the fixing step length of initial value is $0.05s$. Time series to the branch x with $70s$ is produced by the Runge-Kutta algorithms and the total data is 1200. The embedded dimension of the sampling chaotic time series m is 8 by the G- P algorithms. The delay time is $\tau=1$ by the self-correlation function algorithms and the input dimension of the adaptive RBF neural network filtering is 8.The former 1200 data is trained and other 200 data is predicted by the adaptive RBF neural network filtering predictive model.

4.2 Evaluation of the Predictive Ability

The model's predictive ability is generally measure the following three indicators: of MAPE (mean absolute percentage error), RMSE (root mean square error) and RMSPE (root mean square percentage error), they are calculated as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \times 100 \right| \quad (5)$$

$$RMSPE = 100 \times \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{y}_i - y_i}{y_i} \right)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

where, \hat{y}_i is predictive value of the model; y_i is the real value; n is prediction phases, and MAPE assess the predictive capability are as follows: less than or equal to 10% ,then predictive ability is excellent; 10% -20% , then the predictive ability is excellent; 20% -50%, more than 50%, then the prediction is inaccurate. For RMSPE, the prediction square vulnerable to the impact of outliers, for the larger error given greater weight, but still can be modeled on the MAPE to determine the model of the pros and cons. RMSPE values range from zero to infinity. MAPE and RMSPE is the relative indicator, RMSE is the absolute indicator. The RMSE is the smaller, the model predictive ability is the stronger.

4.3 The Simulation Results

That the experimental outcome of Lorenz chaotic sampling time series, the true value (real line) and the predictive value (star line) and the predictive error curve are showed in Fig. 2. Fig. 3. and Fig. 4.

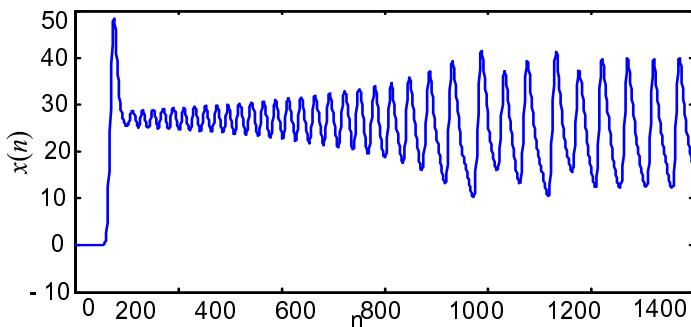


Fig. 2. Lorenz chaotic sampling time series

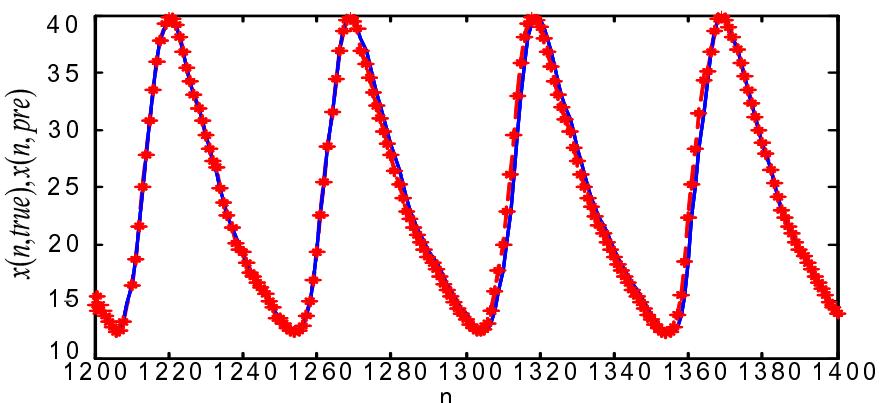


Fig. 3. True value (real line) and predictive value (star line)

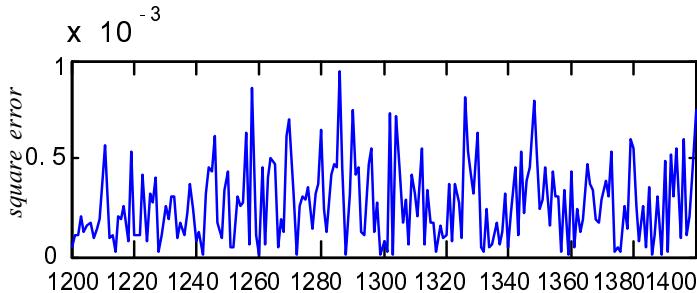


Fig. 4. Predictive error curve

In Fig.2 the sampling chaotic time series number is 1200 by the Runge-Kutta algorithms.

The former 1200 datum are used to learn and train the adaptive wavelet neural networks every 8 datum. After the learned and trained stage, the true value (real line) and predictive value (star line) are shown in Fig.3.

The predictive error curve of the true value and the predictive value is very small in Fig.4.

The true value and the predictive value in the adaptive RBF neural network filtering is to find a inner law in the series itself, which can avoid the disturbance of some subjective factors and enjoys higher reliability. In this study, the fusion of chaotic theory with the adaptive RBF neural network filtering based on chaotic algorithm provides a new method for chaotic time series prediction. The experimental indicated that the network system can enhance the stabilization and associative memory of chaotic dynamics and generalization ability of predictive model even by imperfect and variation inputs during the learning and prediction process by selecting the suitable nonlinear feedback term. Simulation results for the modeling and prediction of chaotic time series show better predictive effectiveness and reliability.

Table 1. Predictive performance comparison table

comparative indicators	BP neural network	RBF neural network
MAPE	5.01%	3.71%
RMSPE	6.13%	4.55%
RMSE	62.50	46.37

From Table 1, the mean absolute percentage error of Lorenz chaotic sampling time series prediction and actual values, BP neural network based on the learning rate variable training algorithm, RBF network based on fast learning algorithm, are 5.1% and 3.71%, respectively. Similarly, for the RMSPE, the results were 6.13% and 4.55%; For RMSE, the results were 62.50 and 46.37. Can be seen from the data on Lorenz chaotic sampling time series RBF network prediction is better than BP neural network.

5 Conclusions

In the paper the chaotic time series RBF neural network model was designed. A RBF neural network Adaptive learning algorithm based on Chaos mechanism was proposed. The method of model selection and algorithm design, are considered the chaos of Lorenz chaotic sampling time series, which is a theoretical value. Simulation results show that the method can reduce MAPE, RMSPE, RMSE, and improve the forecast accuracy, and show better predictive effectiveness and reliability.

Acknowledgement. This research is financially supported by the National Natural Science Funds of China for Distinguished Young Scholar under Grant(50925727), and the Fundamental Research Funds for the Central Universities, Hefei University of Technology for Professor He Yigang, the National Natural Science Foundation of China(NSFC) for Professor Xue-ping Dong (No. 60974022) and the Universities Natural Science Foundation of Anhui Province (No. KJ2012A219) for Professor Yin Lisheng.

References

1. Xue, J., Shi, Z.: Short-Time Traffic Flow Prediction Based on Chaos Time Series Theory. *Journal of Transportation Systems Engineering and Information Technology* 8(5), 68–72 (2008)
2. Gautam, A.K., Chelani, A.B., Jaina, V.K., Devotta, S.: A new scheme to predict chaotic time series of air pollutant concentrations using artificial neural network and nearest neighbor searching. *Atmospheric Environment* 42(18), 4409–4417 (2008)
3. Xiao, Z., Ye, S.-J., Zhong, B., Sun, C.-X.: BP neural network with rough set for short term load forecasting. *Expert Systems with Applications* 36(1), 276–279 (2009)
4. Zhang, H., Changc, M., Wangc, J., Yea, S.: Evaluation of peach quality indices using an electronic nose by MLR. *QPST and BPnetwork* 134(1), 332–338 (2008)
5. Yu, S., Zhu, K., Diao, F.: A dynamic all parameters adaptive BP neural networks model and its application on oil reservoir prediction. *Applied Mathematics and Computation* 195(1), 66–75 (2008)
6. Bakker, R., Schouten, J.C., Giles, C.L., et al.: Learning chaotic attractors by neural networks. *Neural Comput.* 12, 2355–2383 (2000)
7. Leung, H., Lo, T., Wang, S.: Prediction of noisy chaotic time series using an optimal radial basis function neural network. *IEEE Trans. Neural Networks* 12, 1163–1172 (2001)
8. Zhang, Y., Zhou, Q., Sun, C., Lei, S., Liu, Y., Song, Y.: RBF Neural Network and ANFIS-Based Short-Term Load Forecasting Approach in Real-Time Price Environment. *Power Systems* 23(3), 853–858 (2008)
9. Wang, Z., Liu, Y., Liu, X.: State estimation for jumping recurrent neural networks with discrete and distributed delays. *Neural Networks* 22(1), 41–48 (2009)
10. Talebi, H.A., Khorasani, K., Tafazoli, S.: A Recurrent Neural-Network-Based Sensor and Actuator Fault Detection and Isolation for Nonlinear Systems With Application to the Satellite's Attitude Control Subsystem. *Neural Networks* 20(1), 45–60 (2009)
11. Han, M., Xi, J., Xu, S., Yin, F.: Prediction of time series based on the recurrent predictor neural network. *IEEE Transactions on Signal Processing* 52(12) (December 2004)

12. Chiewchanwattana, S., Lursinsap, C., Chu, C.-H.: Time-series data prediction based on reconstruction of missing samples and selective ensembling of FIR neural networks. In: Proceeding of the 9th International Conference on Neural Information Processing, vol. 5 (2002)
13. Wu, J.-D., Hsu, C.-C., Wu, G.: Fault gear identification and classification using discrete wavelet transform and adaptive neuro-fuzzy inference. *Expert Systems with Applications* 36, 6244–6255 (2009)
14. Lee, J.J., Kim, D., Chang, S.K.: An improved application technique of the adaptive probabilistic neural network for predicting concrete strength. *Computational Materials Science* 44, 988–998 (2009)
15. Hu, X.-J., Wang, W., Sheng, H.: Urban Traffic Flow Prediction with Variable Cell Transmission Model. *Journal of Transportation Systems Engineering and Information Technology* 4, 17–22 (2010)
16. Li, K., Chen, T.: Chaotic neural network for nonlinear time series predictions. *Commun. Theor. Phys.* 35, 759–762 (2001)

Estimation of Leaf Weight Based on Snowflake Theory

Lihui Zhou and Hong Wang

College of Sciences, Hebei United University, Tangshan, Hebei, China
zhoulh324@163.com

Abstract. There are two parts in this paper: The first part is the classification of leaf shapes. Leaf shapes are classified from the macro and micro perspectives respectively. In the macro perspective, influential factors on leaf shapes, such as ground diameter, breast diameter, etc., are used as variables to carry out K-means clustering analysis for classification on leaf shapes. In the micro perspective, 40 influential leaf structure factors on leaf shapes are extracted and analyzed by factor analysis and a second K-means clustering. After comparing clustering result with actual classification result, misjudgment probability is found to be very low. In the second part, snowflake model theory is proposed; the growth process of a tree is simulated. Through statistics, the number of leaves growing on smallest branches is not so different to each other. After calculating the number of smallest branches through programming, the total number of leaves could be calculated out.

Keywords: snowflake theory, K-means clustering, factor analysis, error evaluation.

1 Introduction

There is a variety of shapes for leaves. Leaves are the largest organ of trees exposed to air, with the largest contact area to outside environment. Therefore, environmental conditions have a significant impact on shape and structure of leaves. In the evolutionary process trees adapting to different ecological environment, a variety of ecological types of leaves is shaped.

In 2003, BP neural network method has been used by HeShu to solve the problem of leaf shape classification, resulting in an accuracy of 86.67% [1]. However, mathematical model is seldom used to in-depth research of leaf shape classification.

2 Model 1

For trees, there are internal and external causes affecting their leaves shape, but the internal and external causes all have a variety of factors, such as for internal causes there are genes, ways of transportation, and mutation, etc.; for external causes there are sunshine, moisture, temperature, change of worms, and soil etc. Therefore, classification for leaves shapes is a complex and delicate job. Our analysis is mainly carried out from two perspectives, i.e., macro and micro perspectives.

2.1 The Classification Model of Leaves 1

The theoretical result shows that the shape of leaves is not only determined by their growth genes but also affected by growth environment, growth shape and growth scale of the trees. From this perspective, certain influential factors of the shape of tree leaves could be chosen as the indexes. According to relevant material, factors describing shapes of trees include: ground diameter, breast diameter, tree height, clear height, average crown diameter, south-north crown length, east-west crown length, layers, internodes spacing, etc. According to the nine factors cluster analysis is made on trees to classify the similar growth shapes into one category. But it just makes a rough analysis on leaves shapes, so the next step is refined analysis.

2.2 The Classification Model of Leaves 2

Then factor analysis is made on tree leaf shapes within one category to calculate factor score, which is used for clustering. This kind of clustering analysis method is refined. We know that there are several dozens of factors describing leaf shapes, such as leaf shape, leaf width, leaf length, leaf vein, etc., but we know that the length of veins in a certain extent determines leaf length and leaf width. And some factors could be completely described by other factors, so we use the method of reducing dimension firstly and then clustering. We use factor analysis to reduce the dimension of influential factors to get factor score for clustering. This method not only can distinguish well leaf shapes, but also can reduce the complexity of the analyzed problem.

The mathematical model for factor analysis

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + \dots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + \dots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_p = a_{p1}F_1 + a_{p2}F_2 + \dots + a_{pm}F_m + \varepsilon_p \end{cases}$$

Represented with matrix:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}$$

Simply recorded as:

$$X = AF + \varepsilon \quad (1)$$

And meet: (1) $m \leq P$

$$(2) \text{ cov}(F, \varepsilon) = 0$$

$$(3) D(F) = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = I_m$$

F_1, \dots, F_m unrelated and variance are 1

$$(4) D(\varepsilon) = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_p^2 \end{bmatrix}$$

$\varepsilon_1, \dots, \varepsilon_p$ unrelated and different variance

Among them is the P dimensional random vector as unobservable volume, comprised by P indexes got in actual observation. $F = (F_1, \dots, F_m)$ is called common factor of $X = (X_1, \dots, X_p)$ the above-mentioned integrated variable. A is factor loading matrix, on which maximum variance rotation is made with variance, so that the structure of A simplified. In other words, the square value of every column elements of loading matrix is made to polarization 0 or 1 or the more dispersed the contribution rate of public factor is the better is the result. Variables got from factor analysis are represented as linear combination of public factors:

$$X_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + \varepsilon_i \quad i = 1, 2, \dots, P$$

But usually when public factors are used to represent the original variables, it is more convenient to describe the characteristics of research object. Therefore, public factors are represented as linear combination of variables, i.e., the factor score function, namely

$$F'_j = \beta_{j1}X_1 + \beta_{j2}X_2 + \cdots + \beta_{jp}X_p \quad j = 1, 2, \dots, m \quad (2)$$

We calculated m factor score for each left samples. Use the score of these m factors as a variable value to cluster different leaves with the method of K-means Cluster.

2.3 Clustering Error Estimation

We have given the evaluation method for judging clustering effect. Usually we use back substitution misjudgment probability and cross misjudgment probability. If the number of misjudging samples belong to G_1 as belong to G_2 is N_1 , and the number of misjudging samples belong to G_2 as belong to G_1 is N_2 , the total number of samples of the two general classifications is n , Then misjudgment probability is:

$$p = \frac{N_1 + N_2}{n} \quad (3)$$

Back substitution misjudgment probability

Set G_1 , G_2 as two general classifications, X_1, \dots, X_m and Y_1, \dots, Y_n are training samples from G_1 , G_2 respectively, with all the training samples used as $m+n$ new samples, which is substituted gradually into established criterion for judging the ownership of the new samples. The process is called back substitution. If the number of misjudging samples belong to G_1 as belong to G_2 is N_1 , and the number of misjudging samples belong to G_2 as belong to G_1 is N_2 , then misjudgment probability is $\hat{p} = \frac{N_1 + N_2}{m+n}$: Cross judgment probability

Back to generation misjudgment probability is to eliminate a sample every time, and use the rest of $m+n-1$ training samples to establish a criterion for judgment, then use established criterion to make judgment on deleted samples. The above-mentioned analysis is made on each sample of those training samples, and uses its misjudgment proportion as the misjudgment probability. The specific procedure is as follows:

From training samples in general classification G_1 , eliminate one of the samples, and use the rest of the samples $m-1$ plus all samples in G_2 to establish discriminant function;

Use the established discriminant function to make judgment on eliminated samples;

Repeat steps (1), (2) until the samples in G_1 in turn be deleted and judged. The number of misjudged samples is recorded as m_{12} ;

Repeat steps (1), (2), (3) for samples in G_2 , until all of the samples in G_2 in turn be deleted and discriminated. The number of misjudged samples is recorded as n_{21} .

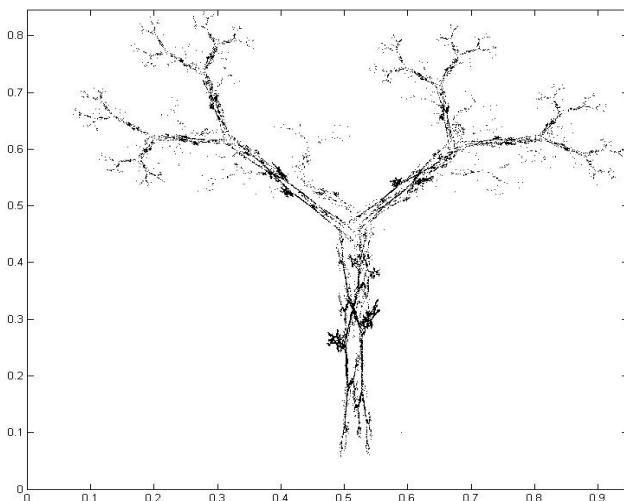
So cross misjudgment probability is estimated: $\hat{p} = \frac{m_{12} + n_{21}}{m+n}$

If clustering result is bad, the following several aspects of optimization could be carried out. 1) Increase sample capacity; 2) Increase new index variables; 3) If statistical data is wrong, rediscover data.

3 Model 2

3.1 Snowflake Theory

Each snowflake on the whole is a hexagonal star, in which there are six trunks, and then each trunk has small branches, and smaller branches growing on small branches, and so on, as shown in figure 1 below. The process of shaping snowflake is copying part and the whole sections of it constantly. The process with the above mentioned of growth characteristics is called snowflake theory.

**Fig. 1.** Snowflake**Fig. 2.** The Tree of Computer Simulation

We already know in the above that each tree species has its own particular branching angle. We think of tree trunk as straight, and from another perspective, we could see it as a lateral branch. We all know that each lateral branch has the function of branching, and all of the lateral branches have the same status. Each layer of the branches will branch in accordance with certain similar rule. According to this growth rule, we simulate the outline of a tree, as shown in Fig.2.

According to the ideas of snowflake theory, the growth process of trees is established until it reaches the state of the tree for observation. The laws of changing between the state of a certain level of branching and the state of its sub-level of branching should be found out to for the recursion relationship of programming.

Among a certain level of branch the main parameters are the quantity of branches, number of sections, interval of sections, azimuth, included angle of branching, curvature, length of branches, and stem. In [1] three ways of branching have been mentioned, i.e., single axis branching, false binary branching and merging axis branching. To simulate the growth of a tree, which way of branching it belongs to should be found. Then after finding out the law of its branching, computer could be used to simulate out its growth process.

3.2 Establishment of Models

From a macro point of View 1, trees have one thing in common in the composition of it's shape and structure, namely the basic constructing element of trees are trunks, branches and leaves. The structuring of each basic element is following a same way: the trunk gives birth to the first layer of branches, which in turn gives birth to the second layer of branches, and so on. The process of giving birth eventually comes to leaves. In the occurrence and development process of the shape of the trees, organizations similar to the existing organizations are constantly copied and added to the existing ones.

Based on the above-mentioned cloning process, eight basic parameters are used to be defined the structure of branches.

Such as in table 1

Table 1. Eight basic parameters

Layer	Layers of tree	[2,8]
H	Height of branches	[0,0,1,0]
R	Bottom radius of branches	[0,0,1,0]
Alfa	Branching angle	[0,90 ⁰]
K	Rattion of top and bottom radius	[0,0,1,0]
P	Height of branching point	[0,0,1,0]
Q	Attenuation of thickness of branches	[0,0,1,0]
M	Attenuation of lenght of branches	[0,0,1,0]

Because of the influence from many kinds of factors such as gravity, wind and sunshine, etc., In the process of their growth, the growth shape of trees in nature has got great uncertainty and randomness. In order to describe shapes of trees more vividly, in the process of establishing mathematical model stochastic function is introduced. Following is a maple tree simulated with a computer model, as shown in Fig.3, Fig.4:

From the simulation rendering with computer, we can find out that the similarity degree between the simulated image and maple tree in real life is very high. Visibly, the reliability of estimating leaves number with the use of snowflake theory is very high.

**Fig. 3.** Simulation of maple**Fig. 4.** Simulation of maple tree

Acknowledgements. The funding organizations are Hebei Higher Social Science Research 2011 Annual Fund(No.SZ2011518) and Tangshan Municipal Bureau of Science and Technology research and development guide plan(No.111302013b), (No.111102033b).

References

1. He, S.: Application of SOM Neural Network on leaves' shape classification. Computer Development and Application 17, Period 2 (2003)
2. Zhou, L., Wang, H., Du, L.: A Balanced Relationship Analysis Between Chinese Economic Growth and the Iron and Steel Production Based on Time Series. In: International Conference on E-Business and E-Government, vol. 2, pp. 192–196 (2010)
3. Jiang, Z.: Study on simulating & virtual growth modeling of *Sabian vulgaris* (TP391.9)
4. Xia, S., Zhu, X.: Review of the measuring method in leaf area. Forestry Survey and Design, Period 2 (2009)
5. Zhang, C., Guo, J., Chen, X.: Research on random walk rough matching algorithm of attribute sub-graph. In: 2011 International Conference on Advanced Materials and Computer Science, pp. 297–302 (2011)
6. Liu, C., Meng, X., Zhang, H.: Research of binocular positioning mathematical model based on the least square method. In: International Conference on Intelligent Structure and Vibration Control, pp. 473–477 (2011)
7. Jiang, Y., Zang, R.: A preliminary analysis on elementary architecture of tropical trees in the topical arboretum of Jian Feng ling, Hainan Island. Resource Science 21, Period 4 (1997)
8. Yang, G., Yili, H.: Mathematical Modeling. China Science and Technology University Press (2008)
9. Liu, F., Zhang, C.: λ -Operations on Packet Sets and the Significance of Application. Journal of Convergence Information Technology (accepted, 2012)
10. Zhou, L.: Analysis about the Effectiveness Evaluation of China's Real Estate Enterprises based on DEA model. In: International Conference on Engineering and Business Management, vol. 1, pp. 1384–1387 (2010)

Efficient Decision-Making Scheme Based on LIOWAD

Meihua Cheng and Shouzhen Zeng

College of Statistics and Mathematics, Zhejiang Gongshang University,
Hangzhou, 310018, China
cmh@tjjj.hz.gov.cn, zszzx1@163.com

Abstract. A new decision making method called linguistic induced ordered weighted averaging distance (LIOWAD) operator by using induced aggregation operators and linguistic information in the Hamming distance. This aggregation operator provides a parameterized family of linguistic aggregation operators that includes the maximum distance, the minimum distance, the linguistic normalized Hamming distance, the linguistic weighted Hamming distance and the linguistic ordered weighted averaging distance, among others. So give special attention to the analysis of different particular types of LIOWAD operators. End the paper with an application of the new approach in a decision making problem about selection of investments under linguistic environment.

Keywords: Linguistic information, IOWA operator, distance measure, decision making.

1 Introduction

The induced ordered weighted averaging (IOWA) operator introduced by Yager and Filev [1] is an extension of the ordered weighted averaging (OWA) operator [2,3]. The difference is that the reordering step is not developed with the values of the arguments but could be induced by another mechanism such that the ordered position of the arguments depends upon the values of their associated order-inducing variables.

Recently, on the basis of the idea of the IOWA operator, Merigó and Casanovas [4] presented an induced ordered weighted averaging distance (IOWAD) operator that extends the OWA operator by using distance measures and a reordering of arguments that depends on order-inducing variables. The IOWAD generalizes the ordered weighted averaging distance (OWAD) operator [5], and provides a parameterized family of distance aggregation operators between the maximum and the minimum distance based on a complex reordering process that reflects the complex attitudinal character of the decision-maker. The IOWAD operator is very useful for decision-making problems because it can establish a comparison between an ideal alternative and available options in order to find the optimal choice. As such, the optimal choice is the alternative closest to the ideal one. They also used the IOWA operator in the Euclidean distance [6]. Then Zeng and Su [7] used the OWA in intuitionistic fuzzy

sets, and developed an intuitionistic fuzzy ordered weighted distance (IFOWD) operator.

When using the IOWAD operator, it is assumed that the available information includes exact numbers or singletons. However, this may not be the real situation found in the decision-making problem. Sometimes the available information is vague or imprecise and it is not possible to analyze it with exact numbers. In this case, a better approach may be the use of linguistic variables. The use of the fuzzy linguistic approach [8] provides a direct way to manage the uncertainty and model the linguistic assessments by means of linguistic variables. The objective of this paper is to develop the linguistic induced ordered weighted averaging distance (LIOWAD) operator by using the IOWA operator and linguistic information in decision making problems with distance measures. Thus, we will be able to provide a model that is able to assess the information in situations with high degree of uncertainty by using linguistic variables. In addition, this will enable us to formulate a more general model by using order-inducing variables in the reordering process. The main advantage of the LIOWAD operator is that it gives a more complete view of the decision problem and then, it is easier to select the alternative that is more in accordance with our interests. We will also develop an application of the new approach in a decision making problem selection of investments under linguistic environment.

2 Paper Preparation

In this section, we briefly review the linguistic approach and the IOWAD.

2.1 The Linguistic Approach

Many problems of the real world cannot be assessed in a quantitative form. Instead, it is possible to use a qualitative one, i.e., with vague or imprecise knowledge that uses linguistic assessments instead of numerical values [8,9]. We have to select the appropriate linguistic descriptors for the term set and their semantics. For example, a set of seven terms S could be given as follows:

$$\begin{aligned} S = \{ & s_{-4} = \text{extremely poor}; s_{-3} = \text{very poor}; s_{-2} = \text{poor}; \\ & s_{-1} = \text{slightly poor}; s_0 = \text{fair}; s_1 = \text{slightly good}; \\ & s_2 = \text{good}; s_3 = \text{very good}; s_4 = \text{extremely good} \} \end{aligned} \quad (1)$$

In order to preserve all the given information, we extend the discrete term set S to a continuous term set $\bar{S} = \{s_\alpha | s_i \leq s_\alpha \leq s_t, \alpha \in [1, t]\}$, where, if $s_\alpha \in S$, then we call s_α the original term, otherwise, we call s_α the virtual term. Therefore, we are able to compute words directly without losing information in the computation process.

2.2 The IOWAD Operator

The IOWAD operator is a distance measure that uses the IOWA operator in the normalization process of the Hamming distance. For two sets $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ and $\beta = (\beta_1, \beta_2, \dots, \beta_n)$, the IOWAD operator can be defined as follows:

Definition 1. An IOWAD operator of dimension n is a mapping IOWAD: $R^n \times R^n \times R^n \rightarrow R$ that has an associated weighting W with $w_j \in [0,1]$ and

$$\sum_{j=1}^n w_j = 1 \text{ such that:}$$

$$f((u_1, a_1, b_1), (u_2, a_2, b_2), \dots, (u_n, a_n, b_n)) = \sum_{j=1}^n w_j d_j \quad (2)$$

where d_j is the $|a_i - b_i|$ value of the IOWAD triplet (u_i, a_i, b_i) having the j th largest u_i , u_i is the order inducing variable, $|a_i - b_i|$ is the argument variable represented in the form of individual distances.

When using the IOWAD operator, it is assumed that the available information includes exact numbers or singletons. However, this may not be the real situation found in the decision-making problem. Sometimes the available information is vague or imprecise and it is not possible to analyze it with exact numbers. In the following, we shall develop the linguistic induced ordered weighted averaging distance (LIOWAD) operator.

3 Linguistic Induced Ordered Weighted Averaging Distance (LIOWAD) Operator

Lets $s_\alpha, s_\beta \in \bar{S}$ be two linguistic variables, Xu [10, 11] defined the linguistic Hamming distance (LHD) between s_α and s_β as follows:

$$d_{LHD}(s_\alpha, s_\beta) = s_\alpha - s_\beta = \frac{|\alpha - \beta|}{2t} \quad (3)$$

where $2t$ is the number of linguistic terms in the set \bar{S} .

Based on above information, we can define a linguistic induced ordered weighted averaging distance (LIOWAD) operator as following:

Definition 2. Let $A = (s_{\alpha_1}, s_{\alpha_2}, \dots, s_{\alpha_n})$ and $B = (s_{\beta_1}, s_{\beta_2}, \dots, s_{\beta_n})$ be two collections of linguistic labels, a LIOWAD operator of dimension is a mapping LIOWAD: $R^n \times \bar{S}^n \times \bar{S}^n \rightarrow R$ that has an associated weighting W with $w_j \in [0,1]$ and $\sum_{j=1}^n w_j = 1$, such that:

$$LIOWAD((u_1, s_{\alpha_1}, s_{\beta_1}), \dots, (u_n, s_{\alpha_n}, s_{\beta_n})) = \sum_{j=1}^n w_j d_{LHD}(s_{\alpha_j}, s_{\beta_j}) \quad (4)$$

Where $d_{LHD}(s_{\alpha_j}, s_{\beta_j})$ is the $d_{LHD}(s_{\alpha_i}, s_{\beta_i})$ value of LIOWAD triplet $\langle u_i, s_{\alpha_i}, s_{\beta_i} \rangle$ having the j th largest of u_i , u_i is the order inducing variable and $d_{LHD}(s_{\alpha_j}, s_{\beta_j})$ is the argument variable represented in the form of individual distance.

The LIOWAD operator is an extension of the IOWAD operator for uncertain environments that can be assessed with linguistic labels. Similar to the IOWAD operator, the LIOWAD operator is commutative, monotonic, bounded, idempotent, nonnegative and reflexive. Note that these properties can be proved with a similar method than the IOWAD operator and thus omitted.

4 Families of LIOWAD Operators

By using a different manifestation of the weighting vector, we are able to obtain different types of LIOWAD operators, such as the linguistic normalized Hamming distance (LNHD), the linguistic weighted Hamming distance (LWHD), the linguistic ordered weighted averaging distance (LOWAD) operator, the step-LIOWAD, the window-LIOWAD, the median-LIOWAD, the olympic-LIOWAD and the centered-LIOWAD.

Remark 1. For example, the linguistic maximum distance, the linguistic minimum distance, the step-LIOWAD, the LNHD, the LWHD and the LOWAD are obtained as follows:

The linguistic maximum distance, if $w_p = 1$ and $w_j = 0$ for all $j \neq p$, and $u_p = \text{Max} \left\{ |s_{X_i} - s_{Y_i}| \right\}$.

The linguistic minimum distance if $w_p = 1$ and $w_j = 0$ for all $j \neq p$, and $u_p = \text{Min} \left\{ |s_{X_i} - s_{Y_i}| \right\}$.

More generally, if $w_k = 1$ and $w_j = 0$ for all $j \neq k$, we get the step-LIOWAD Operator.

The LNHD is formed when $w_j = 1/n$ for all j .

The LWHD is obtained when the ordered position of the u_i is the same as $|s_{X_i} - s_{Y_i}|$

The LOWAD is found if the ordered position of the u_i is the same as the ordered position $|s_{X_i} - s_{Y_i}|$ for all i .

Remark 2. Another particular case is the the olympic-LIOWAD. This operator is found when $w_1 = w_n = 0$ and for all others $w_{j*} = 1/(n-2)$. Note that if $n=3$ or $n=4$, the olympic-LIOWAD is transformed in the median-ILOWAD.

Remark 3. Note that it is possible to present a general form of the olympic-LIOWAD operator, considering that $w_j = 0$ for $j = 1, 2, \dots, k, n, n-1, \dots, n-k+1$; and for all others, $w_{j^*} = 1/(n-2k)$ where $k < n/2$.

Remark 4. Additionally, it is also possible to present the contrary case of the general olympic-LIOWAD operator. In this case, $w_j = (1/2k)$ for $j = 1, 2, \dots, k, n, n-1, \dots, n-k+1$; and $w_j = 0$, for all others, where $k < n/2$. Note that if $k = 1$, then we get the contrary case of the median-LIOWAD.

5 Illustrative Example

In this section, we present a new approach based on the LIOWAD operator to decision making with linguistic preference information. The approach can be used in many decision making problems, such as product management, human resource management, the selection of financial products, and so on.

Assume an investment company wants to make an investment the next year and they have five possible investments:

- (1) Invest in a chemical company called A_1 ;
- (2) Invest in a food company called A_2 ;
- (3) Invest in a computer company called A_3 ;
- (4) Invest in a car company called A_4 ;
- (5) Invest in a furniture company called A_5 ;

When analyzing the investments, the experts have considered the following general characteristics:

- (1) C_1 : Benefits in the short term;
- (2) C_2 : Benefits in the mid term;
- (3) C_3 : Benefits in the long term;
- (4) C_4 : Risk of the investment;
- (5) C_5 : Difficulty of the investment;
- (6) C_6 : Other factors.

After careful analysis of these characteristics, the experts have given the following information shown in Table 1.

Table 1. Available information about the investments

	C_1	C_2	C_3	C_4	C_5	C_6
A_1	S ₆	S ₅	S ₅	S ₆	S ₄	S ₃
A_2	S ₅	S ₇	S ₄	S ₆	S ₄	S ₄
A_3	S ₅	S ₄	S ₆	S ₃	S ₆	S ₅
A_4	S ₃	S ₅	S ₄	S ₅	S ₄	S ₃
A_5	S ₄	S ₂	S ₃	S ₇	S ₅	S ₄
A_6	S ₄	S ₃	S ₇	S ₄	S ₂	S ₆

According to their objectives, the enterprise establishes the following ideal candidate shown in Table 2.

Table 2. Ideal strategy

	C_1	C_2	C_3	C_4	C_5	C_6
I	s_6	s_7	s_7	s_7	s_7	s_6

For some situations, we will use a complex reordering process based on order inducing variables. The results of these variables are shown in Table 3.

Table 3. Order-inducing variables

	C_1	C_2	C_3	C_4	C_5	C_6
U	10	12	15	17	24	18

With this information, it is possible to aggregate the available information in order to take a decision. In this example, we will consider the maximum distance, the minimum, the LNHD, the LWHD, the LOWAD and the ILOWAD operator. The optimal choice would be the alternative closest to the ideal. We assume the following weighting vector $W = (0.1, 0.1, 0.1, 0.2, 0.2, 0.3)$. The results are shown in Tables 4.

Table 4. Aggregated Results

	Max	Min	LNHD	LWHD	LOWAD	ILOWAD
A_1	1.5	0	0.92	1.05	0.7	0.75
A_2	1.5	0	0.83	0.9	0.6	0.75
A_3	2	0.5	0.92	0.9	0.75	0.85
A_4	1.5	1	1.3	1.35	1.25	1.35
A_5	2.5	0	1.17	1.1	0.8	1.1

Note that in these cases, the result indicates the distance between the linguistic variables of the candidate and the ideal one. Then, the lowest value in each method is the optimal result. As we can see, depending on the distance aggregation operator used, the optimal choice is different.

If we establish an ordering of the investments, a typical situation if we want to consider more than one alternative, we get the following orderings shown in Table 5.

As we can see, depending on the distance aggregation operator used, the ordering of the strategies is different. Therefore, depending on the distance aggregation operator used, the results may lead to different decisions.

Table 5. Ordering of the Strategies

Max	$A_1 = A_2 = A_4 \succ A_3 \succ A_5$
Min	$A_1 = A_2 = A_5 \succ A_4 \succ A_3$
LNHD	$A_2 \succ A_1 \succ A_3 \succ A_5 \succ A_4$
LWHD	$A_2 = A_3 \succ A_1 \succ A_5 \succ A_4$
LOWAD	$A_2 \succ A_1 \succ A_3 \succ A_5 \succ A_4$
LIOWAD	$A_1 = A_2 \succ A_3 \succ A_5 \succ A_4$

6 Conclusion

We have presented a decision-making approach that uses distance measures and induced aggregation operator under linguistic environment. We have developed a new distance measure, the linguistic induced ordered weighted averaging distance (LIOWAD) operator. It is a new aggregation operator that provides a parameterized family of linguistic aggregation operators. The main advantage of the LIOWAD is that it is able to assess uncertain problems where the available information can not be represented with numerical values but it is possible to use linguistic ones.

We have analyzed an application of the new approach in a decision-making problem regarding the selection of investments. We have seen that this approach gives a more complete information of the decision problem because it is able to consider a wide range of scenarios depending on the interests of the decision-maker. Moreover, by using order-inducing variables, it is possible to consider different scenarios according to complex attitudinal characters. We have also seen that, depending on the particular type of aggregation operator used, the results may lead to different decisions.

In future research, we expect to develop further extensions of the LIOWAD operator by adding new characteristics in the problem such as the use of the Euclidean distance or the Minkowski distance, and applying it to other problems.

References

1. Yager, R.R., Filev, D.P.: Induced ordered weighted averaging operators. *IEEE Transactions on Systems, Man and Cybernetics B* 29, 141–150 (1999)
2. Yager, R.R., Kacprzyk, J.: *The Ordered Weighted Averaging Operators: Theory and Applications*. Kluwer Academic Publishers, Norwell (1997)
3. Yager, R.R.: On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics B* 18, 183–190 (1988)
4. Merigó, J.M., Casanovas, M.: Decision making with distance measures and induced aggregation operators. *Computers & Industrial Engineering* 60, 66–76 (2011)
5. Merigó, J.M., Gil-Lafuente, A.M.: New decision making techniques and their application in the selection of financial products. *Information Sciences* 180, 2085–2094 (2010)

6. Merigó, J.M., Casanovas, M.: Induced aggregation operators in the Euclidean distance and its application in financial decision making. *Expert Systems with Applications* 38, 7603–7608 (2011)
7. Zeng, S.Z., Su, W.H.: Intuitionistic fuzzy ordered weighted distance operator. *Knowledge-Based Systems* 24, 1224–1232 (2011)
8. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Part 1, *Information Sciences* 8, 199–249 (1975); Part 2, *Information Sciences* 8, 301–357 (1975); Part 3, *Information Sciences* 9, 43–80 (1975)
9. Xu, Z.S.: EOWA and EOWG operators for aggregating linguistic labels based on linguistic preference relations. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12, 791–810 (2004)
10. Xu, Z.S.: Deviation measures of linguistic preference relations in group decision making. *Omega* 33, 249–254 (2005)
11. Xu, Z.S.: Fuzzy ordered distance measures. *Fuzzy Optimization and Decision Making* 11, 73–79 (2012)

Analysis of Sensitive Questions of MSM Based on RRT

Wei Li¹, Ge Gao^{1,*}, Yu-hua Ruan², Xiang-yu Chen¹, and Qiao-qiao Du¹

¹ School of Public Health, Medical College of Soochow University, 215123, Suzhou, China

² STD and AIDS Prevention and Control Center, China CDC, 102206, Beijing, China
gaoge@suda.edu.cn

Abstract. In this paper, multiple choices of sensitive questions model on two-stage sampling was employed to carry a survey on MSM in Beijing, and corresponding formulas were applied to calculate the statistical indicators. The results showed that 87.83 percent of MSM didn't have commercial sexual behavior. The anal sex is the main way of sexual behavior which rate is 64.45%. The MSM in Beijing are associated with high risk factors for HIV infection such as high chance of mobility, lack of physical examination, commercial sexual behavior.

Keywords: MSM, HIV, sensitive question, two-stage sampling, RRT.

1 Introduction

Ministry of Health of the People's Republic of China reported that in China 32.5% of the new 48 thousands HIV infected person in 2009 were transmitted by sexual intercourse of the same sex [1]. It is highly important to acquire the character of the sexual behavior of the men who have sex with men when involved preventing the spread of AIDS. Unfortunately, if we try to ask about the sensitive questions directly towards these people, we may get false reply for their worry about the exposure of privacy, thus caused bias [2]. Many of the sensitive questions studies used usual investigation method such as direct information inquiry, their reliability is doubtful. In this case, we should use Randomized Response Technique (RRT) to improve the reliability when involved investigation of sensitive questions[3].In this paper, Formulas for multiple choices of sensitive questions model were used to estimate related statistical indicators. It was the first time that using these RRT on two-stage sampling to carry a survey on the Men who have sex with men in Beijing of China.

2 Methods

2.1 Sample and Procedure

The study population is the men who have sex with men (MSM) aged from 15 to 50 in Beijing in 2010.The proportion of MSM of the male people aged from 15 to 50 in

* Corresponding author.

Beijing is 1.0% [4], thus we calculated the total number of MSM aged from 15 to 50 in Beijing is 57213. We took the 16 districts in Beijing as the first-stage units ($K_1 = 16$), and took the MSM as the second-stage units, suppose the i th first-stage units include K_{i2} second-stage units. The average of second-stage units in each district is 3576 ($\bar{K}_2 = 3576$). We conducted a two-stage sampling toward MSM in Beijing from August 1, 2010 to October 31, 2010. According to the method of estimating sampling size [5], we randomly select 9 districts ($n_1 = 9$) in first-stage and randomly select 620 people from these 9 districts who are MSM in second stage, the i th second stage include n_{i2} MSM, the average of MSM in each district is 69 ($\bar{n}_2 = 69$). We use excel 2007 and SAS9.13 to analyze the data.

2.2 Multiple Choices of Sensitive Questions Model and Its Formulas

Suppose the choices of answer to sensitive question include k ($k=1, 2, \dots, k$) mutual exclusion categories. Then we used the following randomization set: Place several pieces of cards in a box, each of the cards was marked with a number, the proportion of cards marked with 0 to k are $P_0:P_1:P_2:\dots:P_k$ ($P_0+P_1+P_2+\dots+P_k=1$) respectively. Then each participant was asked to draw a card from the box. When the participant selected the card marked with number 0, he should answer the sensitive category which he belongs to, when the participant selected other cards, he should answer the number marked on the card. In this paper, the proportion of cards marked with 0 to k are $P_0=0.5, P_1=P_2=\dots=P_5=0.1 (k=5)$ when asking about the cost of commercial sexual behavior and $P_0=0.4, P_1=P_2=P_3=0.2 (k=3)$ when asking about the way of sexual behavior respectively.

We use π_{ij} to represent the population proportion of participants who belong to the k th category sensitive question in the i th first-stage unit, p_{ij} is the estimated value of π_{ij} . m_{ij} is the frequency of participants who answered the number j , Let λ_{ij} be the population proportion of people who answer number j in i th first-stage unit, Let $\hat{\lambda}_{ij}$ be the estimator value of λ_{ij} . Then:

$$\hat{\lambda}_{ij} = m_{ij} / n_{i2}$$

According to the total probability formulas [6], we got:

$$\lambda_{ij} = \pi_{ij} p_0 + p_j$$

$$p_{ij} = \frac{\hat{\lambda}_{ij} - p_j}{p_0} \quad i = 1, 2, \dots, n_1; \quad j = 1, 2, \dots, k \quad (1)$$

We use π_j to represent the population proportion of the participants who belong to the k th category sensitive question, according to the formula and results given by Jianfeng Wang [5], the estimated value of π_j and its variance are shown to be:

$$p_j = \frac{\sum_{i=1}^{n_1} K_{i2} p_{ij}}{\sum_{i=1}^{n_1} K_{i2}} \quad (2)$$

$$v(p_j) = \frac{s_{1j}^2}{n_1} \left(1 - \frac{n_1}{K_1}\right) + \frac{s_{2j}^2}{n_1 \bar{n}_2} \left(1 - \frac{\bar{n}_2}{\bar{K}_2}\right) \quad (3)$$

Here, s_{1j}^2 and s_{2j}^2 are shown to be:

$$S_{1j}^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(\frac{K_{i2}}{\bar{K}_2} \right)^2 (p_{ij} - p_j)^2 \quad (4)$$

$$S_{2j}^2 = \frac{1}{\sum_{i=1}^{n_1} K_{i2}} \sum_{i=1}^{n_1} K_{i2} p_{ij} (1 - p_{ij}) \quad (5)$$

3 Results

3.1 Baseline Data

Among the 620 participants, only 28.43% of their household registration is in Beijing, which means most MSM in Beijing come from other provinces of China. 83.87% of them have been in Beijing more than two years; Most of the MSM are very young, 58.39% of their ages are below 30 years. 82.26% of them are single, such as unmarried or got divorce. Many of them have good educational background, more than half of them got associate degree or undergraduate diploma or even master degree. 57.04% of them have a relatively higher income level than local average 3000 RMB per month. 26.94% of them carried physical examination on their genital organ once or twice last year and 10.97 % of them carried none.

3.2 Results of Sensitive Questions

We randomly sampled 49 ($n_{i2} = 49$) MSM in the first district, the frequencies of participants who answered the number j ($j=1,2,\dots,5$) in the first district are 5,5,7,6,26 respectively; thus the population proportion of people who answer number j in first first-stage unit are 5/49,5/49,7/49,6/49,26/49 respectively. According to formula 1, The proportion of MSM whose cost of commercial sexual is under 200 RMB in the first district is:

$$p_{i1} = \frac{\hat{\lambda}_{i1} - p_j}{p_0} = \frac{5/49 - 0.1}{0.5} \times 100\% = 0.4082\% \quad (6)$$

Other districts' proportion can be got in the same way (Table 1).

Table 1. The proportion of commercial sexual cost of each category

Sample	$p_{i1}(\%)$	$p_{i2}(\%)$	$p_{i3}(\%)$	$p_{i4}(\%)$	$p_{i5}(\%)$
1	0.4082	0.4082	8.5714	4.4898	86.1224
2	0.5882	0.5882	1.5686	7.4510	87.8431
3	13.3333	13.3333	2.2222	13.3333	91.1111
4	5.0000	5.0000	5.0000	5.0000	90.0000
5	0.5128	0.5128	0.5128	5.6410	92.8205
6	2.2222	2.2222	2.2222	3.5294	89.8039
7	5.0000	13.3333	5.0000	13.3333	71.6667
8	10.7692	3.0769	3.0769	3.0769	80.0000
9	1.4286	22.8571	1.4286	1.4286	87.1429

According to formula 2, we got the population proportion of the participants who belong to the first category sensitive question is:

$$\begin{aligned}
 p_1 &= \frac{\sum_{i=1}^9 K_{i2} p_{i1}}{\sum_{i=1}^9 K_{i2}} \\
 &= \frac{3328 \times 0.4082\% + \dots + 4062 \times 1.4286\%}{(3328 + 10364 + \dots + 4062)} \\
 &= 3.2988\%
 \end{aligned}$$

According to formulas 4 and 5, we got:

$$\begin{aligned}
 S_{11}^2 &= \frac{1}{9-1} \sum_{i=1}^9 \left(\frac{K_{i2}}{\bar{K}_2} \right)^2 (p_{i1} - p_1)^2 \\
 &= \frac{1}{9-1} \left[\left(\frac{3328}{3576} \right)^2 \times (0.0041 - 0.0330)^2 + \dots + \left(\frac{4062}{3576} \right)^2 \times (0.0143 - 0.0330)^2 \right] \\
 &= 0.0031
 \end{aligned}$$

$$\begin{aligned}
 S_{21}^2 &= \frac{1}{\sum_{i=1}^9 K_{i2}} \sum_{i=1}^9 K_{i2} p_{i1} (1 - p_{i1}) \\
 &= \frac{1}{(3328 + \dots + 4062)} \times \left[3328 \times 0.0041 \times (1 - 0.0041) + \dots + 4062 \times 0.0143 \times (1 - 0.0143) \right] \\
 &= 0.0302
 \end{aligned}$$

The variance of p_1 is:

$$\begin{aligned} v(p_1) &= \frac{s_{11}^2}{n_1} \left(1 - \frac{n_1}{K_1}\right) + \frac{s_{21}^2}{n_1 \bar{n}_2} \left(1 - \frac{\bar{n}_2}{\bar{K}_2}\right) \\ &= \frac{0.0031}{9} \left(1 - \frac{9}{16}\right) + \frac{0.0302}{9 \times 69} \left(1 - \frac{69}{3576}\right) \\ &= 0.0002 \end{aligned}$$

So we got the proportion and its 95% confidence limit of MSM whose cost of commercial sexual is under 200 RMB in the first district is:

$$p_1 \pm 1.96 \times \sqrt{v(p_1)} = 0.0055 \sim 0.0604$$

In the same way we can calculate the proportion and 95% confidence limit of other categories of cost of commercial sexual. We found that most of the MSM didn't have commercial sexual behavior, the proportion of these people is 87.83%, and the left 12.17% represents the proportion of MSM who had commercial sexual behavior in the past year. The cost of commercial sexual behavior are mainly 200~399RMB or above 600 RMB once a time, which constitute 4.96% and 5.86% of total MSM and their 95% confidence limit is 13.71% to 23.77% and 9.11% to 24.12% respectively (Table2).

By using formulas 1 to 5, we found the mainly way of sexual behavior is anal sex, accounts for 64.45% of total MSM and its 95% confidence limit is 56.81% to 72.48%; Oral sex accounts for 18.74% and its 95% confidence limit is 13.71% to 23.77% (Table2).

Table 2. The results of sensitive questions

sensitive questions	Proportion or mean	95% confidence limit
Cost of commercial sexual(RMB)		
0 ~	3.29%	0.5%~6.04%
200~	4.96%	0.56%~9.36%
400~	2.6%	0.94%~4.26%
600~	5.86%	3.21%~8.5%
no charge	87.83%	84.41%~91.25%
Way of sexual behavior		
anal sex	64.45%	56.81%~72.48%
oral sex	18.74%	13.71%~23.77%
other	16.62%	9.11%~24.12%

4 Discussion

Sensitive questions are reflections of contradiction during the developing of society. Acquiring the state of sensitive affairs is highly important for a country to establish

public policies, such as AIDS. It is necessary for us to design some feasible methods to acquire accurate information when conducting a survey on sensitive questions. This paper supplies a method and corresponding formulas on multiple choice of sensitive questions model, which have some meaningful in statistic technique.

When using RRT to carry a survey on sensitive questions, the participant just need to answer "Yes" or "No", or answer a mathematic number, no one knows his answer is for sensitive question or for non sensitive question, so there is no need for him to worry about the exposure of privacy, thus we can get the honest answer. It is the best way to get a reliable estimator of population by using RRT [3].

A limitation of this paper is that we didn't carry evaluation for reliability and validity towards this study, reliability and validity are used to evaluate whether a measurement method can reflect the character of objective things reliably and truly. This study is part of the National Natural Science Foundation project of China. Before this study, a series of related methods were studied which were combined several RRT models with complicated sampling technique. Formulas were deduced and all these methods have highly reliability and validity [4,7,8]. Which indicated that our methods and corresponding formulas are feasible . In order to decrease the bias and get accurate data, two points should be pay attention to when conducting the two-stage sampling. One is training investigator so that each investigator has an overall understanding on two-stage sampling; another is that the sample size is feasible to make sure the sample can represent the population reliably.

Presently there are about 5-10million MSM in China, these MSM usually have several sexual partners and have sexual intercourse without protection [9,10,15]. The MSM has become the second high risk group of infection of AIDS in China. The infection rate of AIDS among MSM has increased from 3% to 10% for the past 3years in some city of China [1]. The high incidence of HIV is consistent with the sharp increase in prevalence among MSM which has been observed in a number of Chinese cities [10-12].It is important for us to make feasible measurement to control the infection of AIDS. We selected Beijing as the survey area because Beijing is one of the chief cities where amount of MSM gather there. Our study shows that most MSM in Beijing come from other provinces of China, this means high possibility of large mobility of MSM, which is high risk for the spread of AIDS; Many MSM are very young and single and have high-income, this may be related to their good educational background. It is just because these reasons preventing the spread of AIDS among MSM can largely reduce the burden of disease. There is 26.94 percent of MSM carried physical examination on their genital organ only once or twice last year and 10.97 percent of them even carried none, this reflects many MSM didn't pay enough attention to self health care. Most of the MSM didn't have commercial sexual behavior last year, which means the places where MSM have sexual behavior are not limited to the commercial sites such as commercial bath, hotels and so on. This reminds that we should make full use of other resources like Internet when we take preventive measures among MSM. The mainly way of sexual behavior is anal sex, which is a high risk way for the spread of HIV. Measurements on safe sexual behavior towards MSM, such as encouraging them to use condoms [10,13],or to find a steady going and regular sex partner [14,15], should be enhanced to prevent the AIDS spread in China even in the world.

References

1. Ministry of Health of the People's Republic of China, UNAIDS, WHO. Epidemic Situation and Preventing Progress of AIDS in China in 2009, R. (2010)
2. Roger, T., Ting, Y.: Sensitive Questions in Surveys. *Psychological Bulletin*, 859–883 (2007)
3. Lensvelt-Mulders, G.J.L.M., et al.: Meta-analysis of randomized response research: 35 years of validation studies. *Sociological Methods & Research* 33(3), 319–348 (2005)
4. PengLiu, Gao, G., He, Z.: Two-stage Sampling on Additive Model for Quantitative Sensitive Question Survey and Its Application. *Suzhou University Journal of Medical Science*, 384–387 (2011)
5. Wang, J., Ge, G., Fan, Y., et al.: The estimation of sampling size in multi-stage sampling and its application in medical survey. *Applied Mathematics and Computation*, 239–249 (2006)
6. Su, L.: Advanced Mathematical Statistic. Beijing University Press (2007)
7. Gao, G., Fan, Y.: Stratified cluster sampling and its application on the Simmons RRT model for sensitive question survey. *Chinese Journal of Health Statistics*, 562–565 (2008)
8. Li, W., Gao, G., He, Z.: Statistical Methods of Two-Stage Sampling on Simmons Model for Sensitive Question Survey with and Its Application. *Studies in Mathematical Sciences*, 46–51 (2011)
9. Cao, G., Guan, W.-H., Wu, X.-G., et al.: Study on infection rate of HIV/SYPHILIS among men who have sex with men in a balneary. *Acta Universitatis Medicinalis Nanjing*, 637–640 (2007)
10. Zhang, D., Bi, P., Lv, F., et al.: Changes in HIV prevalence and sexual behavior among men who have sex with men in a northern Chinese city: 2002–2006. *Journal of Infection*, 456–463 (2007)
11. Feng, L., Ding, X., Lu, R., Liu, J., et al.: High HIV prevalence detected in 2006 and 2007 among men who have sex with men in China's largest municipality: an alarming epidemic in Chongqing, China. *J. Acquire Immune Defic. Syndr.*, 79–85 (2006)
12. Hong, F.C., Zhou, H., Cai, Y.M., et al.: Prevalence of syphilis and HIV infections among men who have sex with men from different settings in Shenzhen, China: implications for HIV/STD surveillance. *Sex Transm. Infect.* 42–46 (2009)
13. Mukandavire, Z., Bowa, K., Garira, W.: Modelling circumcision and condom use as HIV/AIDS preventive control strategies. *Mathematical and Computer Modelling*, 1353–1372 (2007)
14. Meyer, M.A., Champion, J.D.: Protective Factors for HIV Infection Among Mexican American Men Who Have Sex With Men. *JANAC*, 53–62 (2010)
15. Lau, J.T.F., Lin, C., Hao, C., et al.: Public health challenges of the emerging HIV epidemic among men who have sex with men in China. *Public Health*, 260–265 (2011)

Cointegration Analysis Based on Linear Error in Variable Model

Mengtian Yue¹ and Huirong Cao²

¹ Management Office of Scientific Research, Langfang Teachers College,
065000 Langfang, China

² College of Mathematics and Information Science, Langfang Teachers College,
065000 Langfang, China
{yuemengtian,huirongcao}@126.com

Abstract. Linear regression coefficients in traditional cointegration analysis are estimated using ordinary least square criterion, when the response variables and input variables have ignorable observation error, using ordinary linear regression model will lead to cointegration regression model is not unique. In view of this situation, linear error in variable regression model instead of ordinary linear regression model is used to estimate cointegration regression model parameters. At last, computer simulation examples and the cointegration of the Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ) verified the correctness of this method.

Keywords: Cointegration Analysis, Linear Error in Variable Model, Linear Model.

1 Introduction

Much of classical econometric theory has been predicated on the assumption that the observed data come from a stationary process, meaning a process whose means and variances are constant over time. This is, however, not the case for most series in practice. Moreover, working with transformed series makes it difficult to interpret the results or impossible to use the model for forecasting. To overcome this dilemma, Engle and Granger (1987) [1] provided a totally new method noted cointegration analysis for analyzing time series. Cointegration analysis is powerful tool for dealing with a non-stationary time series described the long-term equilibrium relationship.

However, Engle and Granger's cointegration analysis is given based on ordinary least squares method assuming that the independent variables have been measured exactly or observed without error, as such, those models account only for errors in the dependent variable. When independent variables have been measured with errors, estimation based on the standard assumption leads to biased and inconsistent estimates, i.e. the parameter estimator values do not converge to the true values[2].

Errors-in-variables (EIV) model is regression model in which the regressors are observed with errors. EIV model can determine unique relationship between response

series and input series. It is frequently used in computer vision, including motion estimation, shape fitting and image reconstruction [3]. The strong sensitivity of classical procedures to seemingly negligible deviations from the distributional assumptions calls for robust alternative [4].

This paper explains how EIV model is implemented in cointegration analysis and surveys the basic results. The first part covers the class of ordinary linear EIV models, which has been studied in detail by Fuller(1987)[5]. The second part focuses on the cointegration analysis new method based on linear EIV model. In the final sections, we give the simulation results, cointegration analysis on Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ) and draw conclusions.

2 The Estimators for Multiple Linear EIV Model

Errors-in-variable model arises from the study of regression models wherein the covariate is measured with error. We consider the multiple linear EIV model

$$\begin{cases} Y_i = \alpha_0 + x_i^T \beta + \varepsilon_i & (i = 1, 2, \dots, n) \\ X_i = x_i + u_i \end{cases} \quad (1)$$

where $\beta = [\beta_1, \dots, \beta_p]^T \in R^p$, α_0 is an intercept term, (X_i^T, Y_i) , $i = 1, 2, \dots, n$. are observations, The p -dimensional vectors $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ are unknown design points, and (ε_i, u_i^T) $i = 1, 2, \dots, n$ are independent distributed model statistical errors.

For estimated regression coefficients $\hat{\beta}_1, \dots, \hat{\beta}_p$ and an estimated intercept $\hat{\alpha}_0$, so the orthogonal least square estimator is given by

$$\frac{1}{n} \sum_{i=1}^n \frac{(Y_i - \hat{\alpha}_0 - X_i^T \hat{\beta}_0)^2}{1 + \|\hat{\beta}_0\|^2} = \min_{\alpha_0 \in R, \beta_0 \in R^p} \frac{1}{n} \sum_{i=1}^n \frac{(Y_i - \alpha_0 - X_i^T \beta_0)^2}{1 + \|\beta_0\|^2} \quad (2)$$

where $\Theta = R^{n+1}$ is parameter space, but when $\Theta \subset R^{n+1}$ and $\beta = (\alpha_0, \beta_0^T)^T \in R^{n+1}$, the following conclusions will be correct. Anderson[6], Gleser [7] and Fuller [5] show that the estimators for multiple linear EIV model is consistency and asymptotic normality.

The parameters in multiple linear EIV model can be solved by TLS problem in numerical calculating method. Firstly, we analyze the TLS problem by making substantial use of the SVD [8].

Theorem 1[9]. (Solution of the basic TLS problem $X\beta \approx y$).

If the singular value decomposition (SVD) of the matrix $M = [A \ b]$ is defined by

$$[X \ y] = U \Sigma V^T \quad (3)$$

then the estimators for the multiple linear EIV model is $\hat{\beta} = [\hat{\alpha}_0; \hat{\beta}_0]$. where

$$U = [u_1, \dots, u_n], u_i \in R^n, U^T U = I_n \quad , \quad V = [v_1, \dots, v_n], v_i \in R^n, V^T V = I_{p+1} \quad , \\ \Sigma = diag(\sigma_1, \dots, \sigma_{p+1}), \sigma_1 \geq \dots \geq \sigma_{p+1} \geq 0 \quad , \\ \hat{\alpha}_0 = \bar{y} - [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_p] \hat{\beta}_0 \quad \text{and} \quad \hat{\beta}_0 = (A^T A - \sigma_{p+1}^2 I_p)^{-1} A^T b \quad .$$

3 The Cointegration Analysis Based on Linear EIV Model

Definition 1: The time series $\{Y_t\}$ and $\{X_t\}$ are co-integrated based on linear error-in-variable model (1) if and only if $\varepsilon_t, u_{t1}, u_{t2}, \dots, u_{tp} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$.

When the time series $\{Y_t\}$ and $\{X_t\}$ have ignorable observation error, using the cointegration analysis based on linear EIV model will lead to cointegration regression model is unique. However, Engle and Granger's cointegration analysis which based on ordinary least squares method will change according to response variables and input variables exchange.

Theorem 2: The residual series $\{or_t\}$ is integrated of order zero $I(0)$.

Proof: According to equation (1), we have $Y_t - \alpha_0 - X_t^T \beta_0 = \varepsilon_t - u_t^T \beta_0$, Since $\varepsilon_t, u_{t1}, u_{t2}, \dots, u_{tp} \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$, so $\varepsilon_t - u_t^T \beta_0 \sim N(0, (1 + \beta_0^T \beta_0) \sigma^2)$ which implies that $or_t = \frac{Y_t - \alpha_0 - X_t^T \beta_0}{\sqrt{1 + \beta_0^T \beta_0}} \sim N(0, \sigma^2)$.

4 Testing Cointegration Based on Linear EIV Model

From Definition 1 and Theorem 2, we can give testing Cointegration based on linear error-in-variable model:

null hypothesis : H_0 : $\{or_t\}$ is non-stationary and alternative hypothesis: H_1 : $\{or_t\}$ is stationary

The steps of cointegration testing are given below.

Step 1. Compute

$$or_t = \frac{Y_t - \hat{\alpha}_0 - \hat{\beta}_1 x_{1t} - \dots - \hat{\beta}_p x_{pt}}{\sqrt{1 + \hat{\beta}_1^2 + \dots + \hat{\beta}_p^2}}, t = 1, 2, \dots, n.$$

Step 2. Obtain $\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$ Where $\hat{\rho}$ and $S(\hat{\rho})$ can be obtained from the following p -order auto-regression model

$$or_t = \varphi_1 or_{t-1} + \cdots + \varphi_p or_{t-p} + \delta_t. \text{ Therefore,}$$

$$or_t - or_{t-1} = (\sum_{i=1}^p \varphi_i - 1) or_{t-1} - (\sum_{i=2}^p \varphi_i - 1)(or_{t-1} - or_{t-2}) - \cdots - \varphi_{p-1}(or_{t-p+1} - or_{t-p}) + \delta_t$$

Let $\rho = \varphi_1 + \varphi_2 + \cdots + \varphi_p - 1$, if $\{or_t\}$ is stationary then $\rho < 0$, otherwise $\rho = 0$. So we can use $\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$ testing the stationary of $\{or_t\}$.

Step 3. Compute the threshold of $\tau = \frac{\hat{\rho}}{S(\hat{\rho})}$ using Monte Carlo method.

5 Simulation and Case Studies

In order to validate the feasibility and practicality of the cointegration analysis based on linear EIV model proposed in this paper, firstly, we validate the correct of our method by simulating two non-stationary time series $\{X_t\}$ and $\{Y_t\}$, then we use the closed prices of Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ) to further illustrate the performance of our new method.

Specific simulation and case studies are the following:

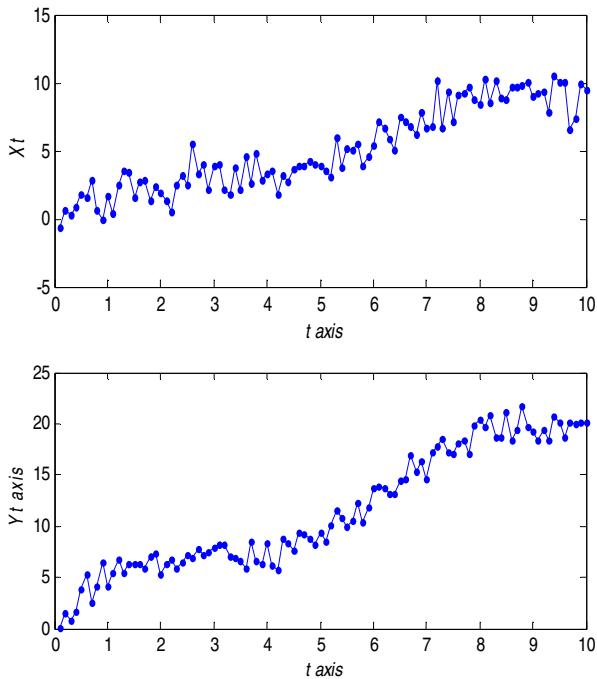
5.1 Simulation Study

The series $\{x_t\}$ and $\{y_t\}$ for the simulation were based on the observations of two time series $\{X_t\}$ and $\{Y_t\}$ respectively. To be specific, define them by

$$\begin{cases} Y_t = \alpha + \beta x_t + \varepsilon_t, \\ X_t = x_t + u_t, \end{cases} t = 1, 2, \dots, n.$$

where $x_t = t + \sin t$, $t = 0.1i + 0.1$ ($i = 0, 1, \dots, 99$), $\alpha = 1$, $\beta = 2$, $y_t = \alpha + \beta x_t$, $\varepsilon_t, u_t \sim N(0, 1)$.

Figure 1 show the $\{X_t\}$ series and $\{Y_t\}$ series are non-stationary. We can also test them by MATLAB. For stochastic simulation of above two series have been measured with errors, any one of them can be response series.

**Fig. 1.** Stochastic simulation of two time series**Table 1.** Parameters estimates of cointegration regression using ordinary linear regression

Input series	Mean of $\hat{\alpha}$	95%confidence interval of $\hat{\alpha}$	Mean of $\hat{\beta}$	95%confidence interval of $\hat{\beta}$
When $\{Y_t\}$ is input series	2.0620	[1.2935,2.8305]	1.7953	[1.6685,1.9221]
When $\{X_t\}$ is input series	0.7139	[1.2935,2.8305]	1.7953	[1.6685,1.9221]

Table 2. Parameters estimates of cointegration regression using linear EIV model.

Input series	Mean of $\hat{\alpha}$	95% confidence interval of $\hat{\alpha}$	Mean of $\hat{\beta}$	95%confidence interval of $\hat{\beta}$
When $\{Y_t\}$ is input series	1.0074	[0.0937,1.9212]	1.9990	[1.8441,2.1539]
When $\{X_t\}$ is input series	1.0074	[0.0937,1.9212]	1.9990	[1.8441,2.1539]

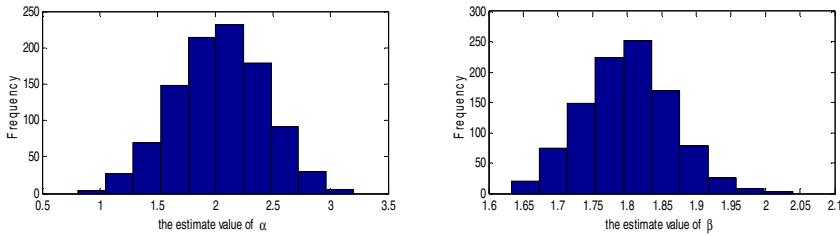


Fig. 2. When $\{Y_t\}$ is response series, the histogram of α and β 's parameter estimates using ordinary linear model.

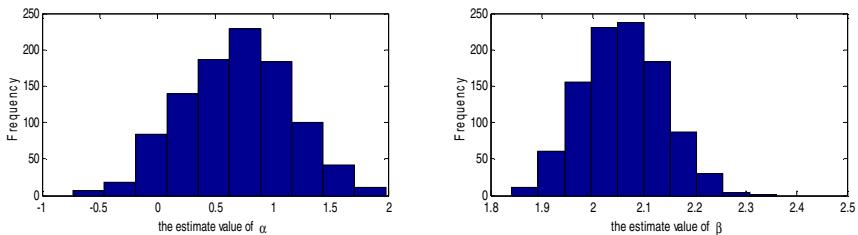


Fig. 3. When $\{X_t\}$ is response series, the histogram of α and β 's parameter estimates using ordinary linear model

The real parameters of linear models are $\alpha=1$ and $\beta=2$. In Table 1 and Figure 2, show that the error between estimate values and real parameters of α and β is too large. Contrast Table 1 with Table 2, we can see the cointegration regression model parameters using EIV models is precise and unique. From Figure 2, Figure 3 and Figure 4, it is easy to see the new cointegration regression model which we presented is precise, too.

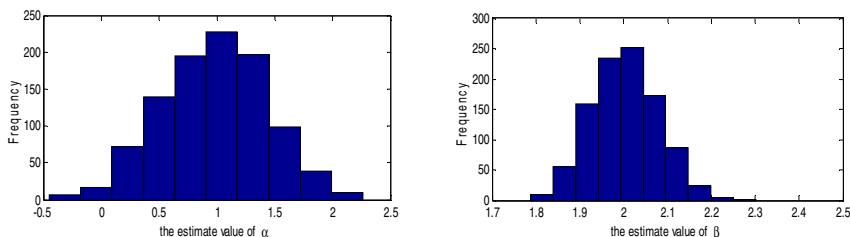


Fig. 4. The histogram of α and β 's parameter estimation using linear EIV model

5.2 Applications in Cointegration Analysis for the Closed Price of Shanghai Stock Index (SZZS) and Shenzhen Component Index (SZCZ)

The 242 trading days' closed prices data of Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ) are from April 29, 2010 to April 29, 2011 on

website <http://stock.sohu.com>. We plot the data in figure 5. The horizontal axis shows the trading days and the vertical axis shows the index values corresponding to SZZS and SZCZ respectively.

Figure 5 show that Shanghai Stock Index (SZZS) and Shenzhen Component Index (SZCZ) are non-stationary. We can also test them by the command dfARTest in MATLAB [10]. Using ordinary linear regression model will lead to cointegration regression model is different.

If the SZCZ is response series, the cointegration regression model based on ordinary linear regression model is

$$SZCZ = -5186.6763 + 6.8039SZZS \quad (4)$$

If the SZZS is response series, the cointegration regression model based on ordinary linear regression model is

$$SZZS = 1254.6026 + 0.1764SZCZ \quad (5)$$

And it can be rewritten as

$$SZCZ = -7111.9663 + 5.36687SZZS \quad (6)$$

If the cointegration regression model based on linear error-in-variable model is

$$SZCZ = -7066.6339 + 6.7875SZZS \quad (7)$$

According to (4) and (6), we can see the cointegration regression model based on ordinary linear regression model is not unique. However, EIV model can give unique relationship between response series and input series.

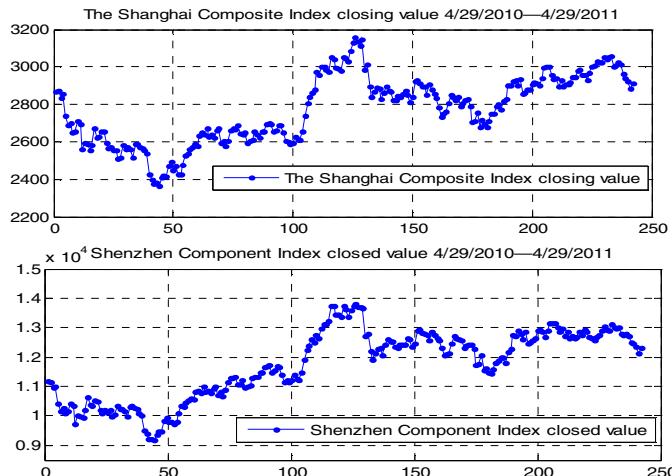


Fig. 5. Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ)

6 Conclusion

We presented a new cointegration analysis method using multiple linear EIV model instead of the traditional linear model. This is a new view of the cointegration analysis, using the model to overcome the shortcomings of traditional. Simulation studies can be seen the new cointegration regression parameters are more precise than the traditional linear model. Finally, we give the results of our algorithm on Shanghai Stock Index(SZZS) and Shenzhen Component Index(SZCZ).It can be seen EIV model can give unique relationship between response series and input series.

Acknowledgments. The work was supported by the Hebei Provincial Science and Technology Agency Soft Science Project (Research on spatial-temporal dynamic evolution and cointegration of Beijing-Tianjin-Hebei Regional Economic system), Hebei Provincial Natural Science Foundation of China(No.A2011408006).

References

1. Engle, R.D., Granger, C.W.J.: Co-integration and error-correction: Representation, estimation and testing. *Econometrica* 55, 251–276 (1987)
2. Tang, S.Z., Li, Y.: The Statistical Basis of Mathematical Models of Biological. Science Press, Beijing (2002) (in Chinese)
3. Matei, B., Meer, P.: A General Method for Errors-in-Variables Problems in Computer Vision. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2000, vol. 2, pp. 18–25 (2000)
4. Choukroun, A., Charvillat, V.: Bucketing Techniques in Robust Regression for Computer Vision. In: Bigun, J., Gustavsson, T. (eds.) SCIA 2003. LNCS, vol. 2749, pp. 609–616. Springer, Heidelberg (2003)
5. Fuller, A.W.: Measurement Error Model. John Wiley, New York (1987)
6. Anderson, T.W.: Estimating linear statistical relationships. *Annals of Statistics* 12(1), 1–45 (1984)
7. Gleser, J.L.: Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. *Contemporary Mathematics* 112, 99–114 (1990)
8. Van Huffel, S., Lemmerling, P.: Total least squares and errors-in-variables modeling: analysis, algorithms and applications. Springer (2002)
9. Wei, M.S.: Generalized Least Square Theory and Computational Problems. Science Press, Beijing (2006) (in Chinese)
10. How do I use AIC and BIC to choose the optimal lags from DFARTEST, DFARDTEST, and DFTSTEST in the GARCH Toolbox?, <http://www.mathworks.cn/support/solutions/en/data/1-72H1G0/index.html?product=ML&solution=1-72H1G0>

Statistical Analysis of Polycyclic Aromatic Hydrocarbons Projects

Jianguo Feng, Guihua Li, Jintang Zhang, and Le Xia

Shandong Provincial Key Laboratory of Depositional Mineralization & Sedimentary Minerals, College of Geological Sciences & Engineering, Shandong University of Science and Technology Qingdao Shandong Province, 266590, China
fengjianguo20316@sohu.com

Abstract. With the rapid economic and social development of China, the attention to environment is higher and higher by common people. To analyses central government's attention level, projects of polycyclic aromatic hydrocarbons (PAHs) funded by the National Natural Science Foundation of China (NSFC) in 1999 ~ 2011 were analyzed from project quantity distribution, fund distribution, subject distribution, units and departments distribution. All the data were collected by search function on ISIS. The results show that: the number of projects funded and funding improved steadily; in the subject, units and departments in terms of uneven distribution, showed strong dispersion; for a long period of time, the question will always be researching focus.

Keywords: National Natural Science Foundation of China, polycyclic aromatic hydrocarbons, statistical analysis.

1 Introduction

The competition among nations is the competition of technology and talent. As reported on the 17th congress of the Communist Party of China, it is the core of the development strategy and the key to improve the comprehensive national strength to build an innovative country. In 21st century, our scientific undertaking is developing by leaps and bounds, which is owed to the effect of the National Natural Science Foundation of China (NSFC) [1].

In recent years, as the steady improvement of the comprehensive national strength, China is paying more and more attention to strengthening the basic research, and the NSFC is one of important projects on basic researches. The NSFC was founded in 1986, which has formed the evaluation principle that relying on experts, promoting democracy, supporting preferentially, equitable and rational, established a mechanism that equal competition, scientific and democratic, encouraging creativity, brought into play the function that guidance, stabilization, incentives. After 20 years of continuous exploration and development, NSFC has formed three funding levels of importance project, emphasis project and surface project .It also has formed a funding pattern which is combined of many kinds of funding types and special funds [2].

Now, the research data about the funding of the NSFC is as follows:

1) The NSFC has published the statistical reports of NSFC and the relevant contents can be got from Statistical projects on the home page of the NSFC (As of press time, it contains the data from 2001 to 2010. <http://www.nsfc.gov.cn/nsfc/cen/xmtj/index.html>). The reports of 2010 analysis from the applications and funding of the surface project, the Outstanding Youth Science Foundation, the key project, the National Science Found for Distinguished Young Scholars of China, the Research Fund for Scholars of Hong Kong, Macao and overseas and so on (the grant of the NSFC in 2010).

2) Some experts and scholars have done some analysis aimed at the NSFC, such as the analysis to the surface project in the field of atmospheric science from 1986 to 2006 using the quantitative data of the project application forms and concluding reports on the project files' electronic inquiring system of the NSFC by Xiaogang Zhou [3], the analysis to the NSFC from 2006 to 2008 through the number and the importance of items by Ying Li [4], the analysis to the NSFC about entomology of 2010 by Xiuxia Lv [5].

Polycyclic Aromatic Hydrocarbons (PAHs) is a kind of hydrocarbons containing two or more benzene rings and lining in line, angle, or clumps. There are more than 400 kinds with the effect of causing cancer in the PAHs which has been found. And [a] has the strongest cancer toxicity, so it is used as the environment pollution index by PAHs. PAHs have become the organic pollutants with the common concern. EPA issued 129 kinds of priority monitoring pollutants in 1979, including 16 kinds of PAHs. China also has listed 7 kinds of PAHs in China environmental priority pollutants black list. The accumulation in the environment of PAHs is a serious threat to human health [6-7].

Zhou Lihong, Wen-Chang Zhao [8-9] have shown that, PAHs in the air, water, river (or lake) sediments and soil, a large part comes from the hydrocarbons in coal incomplete combustion or hydrolyze in a reducing atmosphere. In addition, a part of aromatic compounds forms low molecular weight PAHs in plant coalification process, and this part of PAHs in coal with solid or liquid phase. PAHs in the raw coal were heated and evaporated out is another source of PAHs in the environment [10].

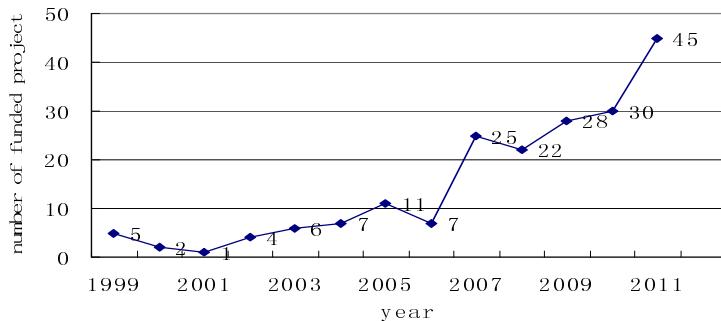
This paper uses the comprehensive search function on ISIS system projects of the NSFC (http://159.226.244.22/portal/proj_search.asp) with PAHs for the project subject. It does an inquiry to the funded projects from 1999 to 2011 and makes a statistical analysis from the NSFA and in China's scientific research development.

2 Funding Statistics

2.1 Number of Funded Project

NSFC has aided 193 projects related to PAHs from 1999 to 2011. Figure1 shows the number of foundation project of each year.

Figure1 shows that the funding number related to PAHs form 1999 to 2006 is relatively less, and rapid increase from 2007, 2011 up to 45 items.

**Fig. 1.** Number of funded project related to PAHs

2.2 Funding

The minimum funding is 80000 RMB, and the maximum is 3.4 million RMB of 193 funding projects in 1999-2011. Table 1 illustrates the funding from 1999 to 2011.

Table 1. Funding statistical table of NSFC on PAHs projects

Year	Units: ten-thousand RMB						
	1999	2000	2001	2002	2003	2004	2005
Funding	100	39	200	89	263	181	337
Year	2006	2007	2008	2009	2010	2011	
Funding	213	804	517	872	918	2266	

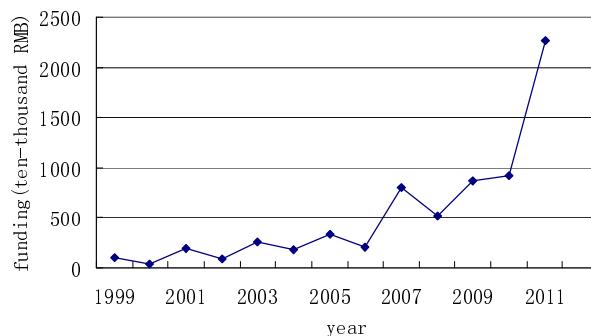
**Fig. 2.** Line graph of funding on the PAHs project

Figure 2 shows that the annual funding has an increasing trend year by year, especially in recent years. There are two reasons promote funding increasing: one is the increasing on the number of funding projects (Figure 1), the other is the increasing of single project. For example, youth foundation is an average of 250000 RMB per project in 2011, which is up by 25% than before.

Table 2. Interval distribution statistical of funding

Funding	1-10	11-20	21-30	31-40	40-80	>81
Project number	4	43	61	36	45	4

Table 2 shows that there are 185 projects between 110 and 800 thousand RMB, that is about accounted for 96% of all founded projects (Not consider the funding difference in different years by economic growth).

2.3 Subject Distribution

The NSFC set following science departments: the mathematical science department(A), the chemistry department(B), the life science department(C), the geosciences department(D), the engineering and material science department(E), the information science department(F), the management science department(G), the medical science department(H), the project agency(J), the United Fund field(L), the office(M), the policy bureau(Z). The Number of foundation project related to PAHs is the most founded by the geosciences department, and the life science department is the second (Table3).

Table 3. Distribution of foundation project related to PAHs of department

Department	B	C	D	E	H
Number	59	17	96	14	7

The geosciences departments include seven subjects: geography (D01), geology (D02), geochemistry (D03), geophysics and space physics (D04), atmospheric sciences (D05), marine sciences (D06), Petroleum chemical mutual funds (D07).

Table 4. Distribution of foundation project related to PAHs of subject

Subject	D01	D02	D03	D05	D06
Number	44	8	21	3	20

Table4 shows that the Number of foundation project related to PAHs is the most founded by geography subject, geochemistry subject and marine sciences subject are the followed.

2.4 Units distribution

There are 82 units were founded during 13 years (1999-2011). 19 unit's foundation project number exceeds 2, 14 unit's foundation project number is 2 and 49 unit is only 1. (Table5)

Table 5. Units of foundation project number exceeds 2

units	Number
Xiamen University	14
Peking University	13
Zhejiang University	11
Nanjing University	8
Nanjing Agricultural University	7
Institute of Applied Ecology, Chinese Academy of Sciences	7
Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences	7
Beijing Normal University	6
Dalian University of Technology	5
South China University of Technology	5
Chinese Research Academy of Environmental Sciences	5
Tsinghua University	4
Guangzhou Institute of Geochemistry, Chinese Academy of Sciences	4
Central South University of Forestry and Technology	4
Fudan University	3
Nankai University	3
Ocean University of China	3
Institute of Geochemistry, Chinese Academy of Sciences	3
Sun Yat-sen University	3

It can be seen from Table5 that the unit mainly include universities and institutes. Table6 shows the statistical result of all the 193 foundation projects by units.

Table 6. Statistical of foundation project by units

	number of units	number of foundation projects
University	59	141
Institutes	23	52

It can be perceived that university is the main part of foundation projects related to PAHs, showing that the function of teaching and scientific research in universities. We should also pay attention to the difference on basic research strength among different universities at the same time.

3 Conclusions

Secretary-General Jintao Hu point out that basic research can lead to the development of science and technology. With intensively basic research being supporter, we can continuously improve our initial innovation capability and persistently national development [11].

Currently, China government and society pay more attention to environmental problems and capital input is increasing continuously. As experience shows, the study group and the funds investment on a question is function of time, with an approximately normal distribution discipline. From the above analysis we can see that the research on "PAHs" will be persistent overheating for a period of time. So the government will increase the foundation for this area.

Acknowledgments. Supported by National Natural Science Foundation of China (40802086).

References

1. Hou, D.: Statistical Analysis of Distribution and Output of NSFC General Projects from 2001 to 2007. *Journal of Modern Information* 28, 197–200 (2008)
2. Chen, L., Li, J., Zheng, S., et al.: Distribution Statistical Research of the Units Founded by NSFC. *Science and Technology Management Research* 28, 106–108 (2008)
3. Zhou, X., Luo, Y.: A Summary on General Program Projects in Atmospheric Science Funded by the NSFC from 1986 to 2006. *Advances in Earth Science* 22, 540–546 (2007)
4. Li, Y., Qiu, G., Meng, Z.: Statistical Analysis of Programs Funded by NSFC. *Management of Agricultural Science and Technology* 28, 25–28 (2009)
5. Lv, X., Qi, J., Zhang, X.: Statistics and Analysis Related to Entomology Funded by NSFC. *Chinese Journal of Applied Entomology* 48, 214–224 (2011)
6. Wei, J., Wu, J., Jiao, W.: Toxicity of Polycyclic Aromatic Hydrocarbons and Its Disposal. *Pollution Control Technology* 21, 65–69 (2008)
7. Janet ·larson writing, Qiuye translate: Cancer Now Leading Cause of Death in China. *World Environment* 4, 41–42(2011)
8. Zhou, L.: Pahs Pollution Source and Prevention and Control Measures. *Rural Economy and Science Technology* 21, 126 (2010)
9. Zhao, W., Cheng, J., Xie, H., et al.: PAHs: Sources, Pathway and Their Monitoring and Analysis. *Environmental Science and Technology* 29, 105–107 (2006)
10. Zhang, H., Liu, G., Xue, J., et al.: Study of Polycyclic Aromatic Hydrocarbons (PAHs) and its environmental Impact in Coal and Coal Combusting Products. *Journal of China Coal Society* 30, 97–101 (2005)
11. Chen, Y.: Studying and Implementing the Scientific Outlook on Development, Contributing to the Sound and Rapid Development of Basic Research in China. *Bulletin of National Natural Science Foundation of China* 23, 1–6 (2009)

Research on CNC Contour Error Compensation Approach Based on Straightway Approximate Nodes

Guoyong Zhao, Zhiyong Li, and Hongjing An

Department of Mechanical Engineering, Shandong University of Technology Zibo 255049,
Shandong Province, China
Zgy709@126.com

Abstract. In the CNC machining on complex parts, how to compute contour error with high precision and distribute contour error correction quantity on each sampling period, is much significant to improve matching degree among each axis. Consequently, a CNC contour error compensation control approach with stable calculation error, high computing precision and satisfied real-time characteristic, is researched in detail in the paper. Firstly, compute the contour error according to the actual cutter position dots and the straightway approximate nodes; Secondly, add the obtained contour error to the following error of current sampling period, and send the results to CNC position controller to calculate position controlled quantity; Finally, the contour error compensation contrast experimentations are done on the three-axis linked CNC test table, and the experimental results show that the developed approach can reduce contour error effectively.

Keywords: Approximate nodes, CNC system, Contour error compensation, Complex parts.

1 Introduction

In general, multi-axis CNC machine tool is adopted to process complex parts, after approximating complex cutter position track instruction curve with straightway. To multi-axis CNC machine tool, the profile precision is the important factor to determine its machining accuracy [1]. Because CNC machine tools have complicated servo drive equipments, and the CNC system parameters may change in practical machining, the multi-axis actual dynamic performances don't match well, which reduces the profile precision [2-3]. In contrast to the advanced single-axis servo controller, the cross-coupled-controller is more effective to enhance profile precision, which computes the contour error and compensates each axis servo motor on each sampling period [4-6].

Some research results in point have been achieved recently. For instance, Myung-Hoon LEE puts forward a multi-axis contour controller based on a contouring error vector using parametric curve interpolation [7]. Peng Chao-Chung introduces a new contour index (CI) aimed to arc and line profile, which can be looked as an equivalent contouring error [8]. Aimed to profile curve in plane and space, Gen Lirong and

Wang Baoren look the distance of actual position to the line which links the dots of the current and the last sampling period as the current contouring error respectively, entitled “two-dot line approximation method” [9-10]. However, because of inertia and frictional force, the hysteresis phenomena exist in truly CNC machine tool each axis movement, which is difficult to be foreseen accurately. As a result, the above-mentioned approaches are unstable sometimes. What's more, the calculation error is uneasy to control if the hysteresis time is much longer than a sampling period.

Consequently, a CNC contour error compensation control approach with stable calculation error, high computing precision and satisfied real-time characteristic is developed in the paper. Above all, compute the contour error according to the actual cutter position dots and the straightway approximate nodes; Secondly, add the obtained contour error to the following error of current sampling period, and send the results to CNC PID position controller to calculate position controlled quantity in order to compensate contour error; Finally, the contour error compensation control contrast experimentations are done on the three-axis linked CNC test table.

2 Contouring Error Computing Model

The key idea of the developed contour error computing model is as followed: After approximating complex parts cutter position track instruction curve with straightway according to equi-error method, calculate the current actual cutter position coordinates owing to the position measure feedback from each axis and table on each CNC line interpolation sampling period; Compute the minimum distance from current actual cutter position to cutter position track instruction curve according to the actual cutter position dots and the approximate nodes, in other words, to calculate the contour error.

As shown in Fig.1, suppose to approximate part cutter track instruction curve L under the precision requirement with straightway AB, BC..., and define the actual cutter position as dot R on certain sampling period. Above all, obtain the three approximate nodes A, B, C nearest to actual cutter position R on the cutter position instruction curve L, and then calculate the distance $|RM|$, $|RN|$ from actual cutter position R to straightway AB, BC. It is noticed that the calculation is complicated if transform the distance from dot to line, to the maximum distance from dot to plane pencil through the line. Consequently, the vector method with the space analytic geometry and vector algebra theory is adopted to compute the distance $|RM|$, $|RN|$ from dot R to straightway AB, BC:

$$|RM| = \frac{|AB \times AR|}{|AB|} \quad (1)$$

$$|RN| = \frac{|BC \times BR|}{|BC|} \quad (2)$$

The coordinates of both approximate nodes A, B, C and actual cutter position R are known, so the calculations of Equation (1) and (2) are simple. After obtaining $|RM|$ and $|RN|$, the contour error \mathcal{E} is calculated according to two kinds of conditions.

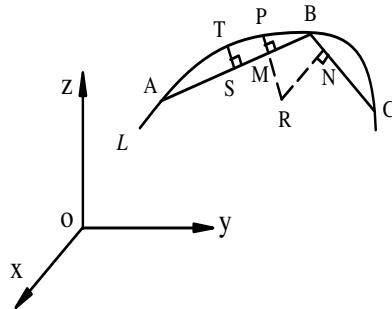


Fig. 1. The CNC contour error computing model

(1) If $|RM| \leq |RN|$, obtain the contour error with Equation (3):

$$\varepsilon \approx RM \quad (3)$$

As shown in Fig.1, the approximate error ST is constant, suppose the intersection point of RM and curve L be dot P, then the calculation error of Equation (3) is MP.

Because $|MP| \leq |ST|$, the calculation error of contour error computing model is less than or equal to approximate error.

(2) If $|RM| > |RN|$, obtain the contour error with Equation (4):

$$\varepsilon \approx RN \quad (4)$$

In like manner, the calculation error of contour error computing model is less than or equal to approximate error.

3 Contouring Error Compensation Approach

Except for the three PID position controller for X axis, Y axis and Z axis, Myung-Hoon LEE sets up an additional PID contour error controller [7]. The calculation approach is rather complicated. In the paper the contour error control compensation approach is developed, which adds the obtained contour error to the following error of current sampling period, and sends the result to CNC PID position controller to calculate position controlled quantity. The CNC contour error calculation and compensation program flow chart based on straightway nodes is shown in Fig.2.

Firstly, after receiving the N^{th} machining program segment coding and pretreatment results on the K^{th} sampling period, interpolate and obtain following error Ex , Ey , Ez ; Secondly, find the three approximate nodes nearest to actual cutter position of current sampling period, and calculate contour error ε with Equation (3) and (4); Thirdly, decompose ε to ε_x , ε_y , ε_z along X, Y, Z coordinate axes, and

compute each axis optimal displacement of current sampling period after contour error compensation, which is μ_x, μ_y, μ_z ; Finally, input the μ_x, μ_y, μ_z to X, Y, Z coordinate axes PID position controller respectively, and compute the correction quantity to control X, Y, Z coordinate axes servo motors.

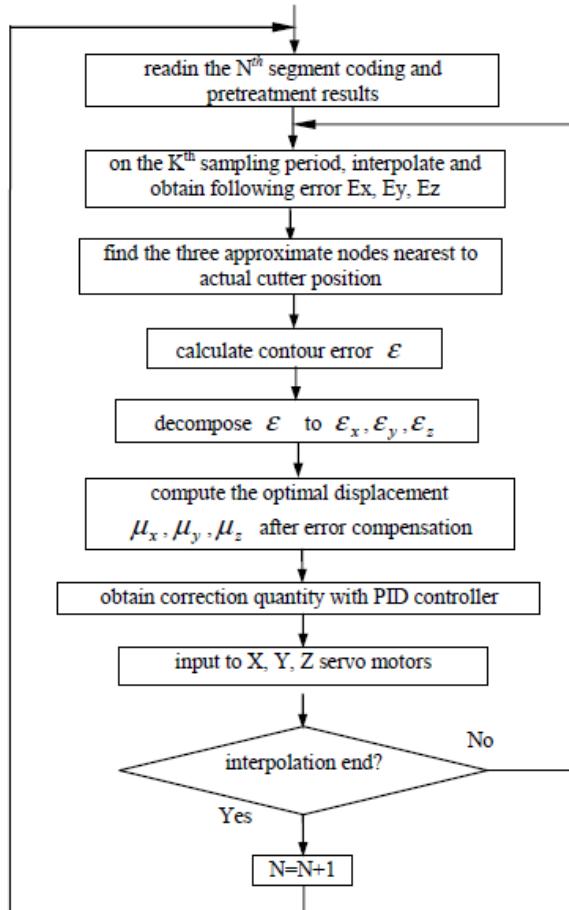


Fig. 2. The CNC contour error calculation and compensation program flow chart

4 Experimentations on Contour Error Compensation Control

The developed three-axis linked table CNC controller is made up of PC and DSP movement control card. The PC and DSP movement control card communicate through USB2.0. The PC acts as the man-machine interface, which implements instruction control, code compilation, states display and other functions; And the interpolation, position control and contour error compensation control function are carried out on the DSP movement control card.

The traditional contour error “tangent approximation method” and the developed contour error compensation control approach based on straightway approximate nodes are programmed on DSP movement control card. Suppose the part cutter track instruction curve be L1, described with NURBS curve: Let the control points be A1(50,0,50), A2(50,50,50), A3(0,50,0), A4(-50, 50,-50), A5(-50, 0,-50), A6(-50, -50,-50), A7(0,-50,0), A8(50,-50,50), A9(50,0,50); the corresponding weights be (1,0.6,1, 0.4,1,0.4,1,0.6,1); the knot vector be (0,0,0,0,0.25,0.375,0.5,0.625, 0.75,1,1,1,1).

When adopting the traditional “tangent approximation method”, the contour error is shown in Fig. 3, with the contour error max. 0.061mm; When approximating complex part cutter instruction curve with straightway according to equi-error method and adopting the introduced approach, the contour error is shown in Fig.4, with the contour error max. 0.039mm. The developed approach reduces the contour error and enhances the profile precision effectively in contrast to Fig. 3.

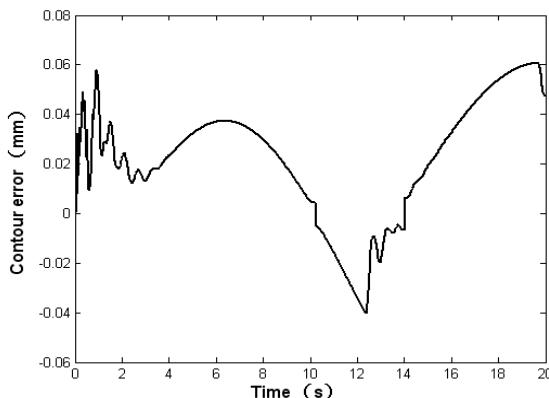


Fig. 3. Contour error adopting the traditional “tangent approximation method”

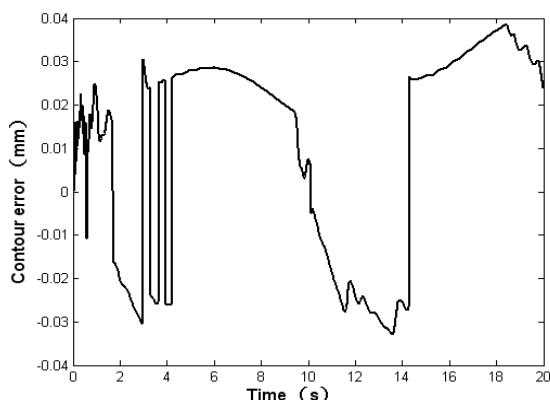


Fig. 4. Contour error adopting the developed approach

5 Conclusions

A CNC contour error compensation approach based on straightway approximate nodes is researched in detail in the paper. Firstly, compute the contour error according to the actual cutter position dots and the straightway approximate nodes; Secondly, add the obtained contour error to the following error of current sampling period, and send the results to CNC PID position controller to calculate position controlled quantity; Finally, the contour error compensation control contrast experimentations are done on the three-axis linked CNC test table. The strongpoints of the developed approach are concluded: For one thing, the calculation error of the introduced contour error computing model is less than or equal to approximate error; For another thing, make use of each axis position controller to compensation contour error and needn't add additional contour error controller, which is helpful to CNC system real-time characteristic.

Acknowledgments. The authors are grateful to the Project of the National Natural Science Foundation of China (No.51105236), and the Shandong Province Promotive research fund for excellent young and middle-aged scientists of China (No. BS2011ZZ014).

References

1. Zhao, X., Guo, Q.: Zero Phase Adaptive Robust Cross Coupling Control for NC Machine Multiple Linked Servo Motor. Proceedings of the CSEE 28(12), 129–133 (2008)
2. Su, K.-H., Cheng, M.-Y.: Contouring accuracy improvement using cross-coupledcontrol and position error Compensator. International Journal of Machine Tools and Manufacture 48(12), 1444–1453 (2008)
3. Cheng, M.-Y., Su, K.-H., Wang, S.-F.: Contour error reduction for free-form contour following tasks of biaxial motion control systems. Robotics and Computer-Integrated Manufacturing 25(2), 323–333 (2009)
4. Liu, Y., Cong, S.: Optimal Contouring Control Based on Task Coordinate Frame and Its Simulation. Journal of System Simulation 21(11), 3381–3386 (2009)
5. Zhong, Q., Shi, Y., Mo, J.: A Linear Cross-Coupled Control System for High-Speed Machining. International Journal of Advanced Manufacturing Technology 19(8), 558–563 (2002)
6. Yeh, S.-S., Hsu, P.-L.: Analysis and Design of Integrated Control for Multi-Axis Motion Systems. IEEE Transactions on Control Systems Technology 11(3), 375–382 (2009)
7. Lee, M.-H., Yang, S.-H., Kim, Y.-S.: A multi-axis contour error controller for free form Curves. JSME International Journal 47(1), 144–149 (2004)
8. Peng, C.-C., Chen, C.-L.: Biaxial contouring control with friction dynamics using a contour index approach. International Journal of Machine Tools and Manufacture 47(10), 1542–1555 (2009)
9. Geng, L., Zhou, K.: Research on Real Time Compensation Method Based on Time Series Predictive Technology for Contour Error of CNC Machine Tool. Manufacturing Technology & Machine Tool 6, 22–25 (2004)
10. Wang, B., Wang, J., Zhang, C.: Contour error vector model and its application to CNC systems. Computer Integrated Manufacturing Systems 16(7), 1401–1407 (2010)

Least Squares η -Hermitian Solution for Quaternion Matrix Equation $AXB = C$

Shifang Yuan

School of Mathematics and Computational Science, Wuyi University,
Jiangmen 529020, P.R. China
ysf301@yahoo.com.cn

Abstract. By applying the methods presented by Yuan et al. [Yuan, S.F., Wang, Q.W.: Two special kinds of least squares solutions for the quaternion matrix equation $AXB + CXD = E$. Electron. J. Linear Algebra 23, pp. 257--274 (2012)], we derive the expression of the least squares η -Hermitian solution of the quaternion matrix equation $AXB = C$ with the least norm.

Keywords: Matrix equation, Least squares solution, Quaternion matrices.

1 Introduction

Let $R^{m \times n}$, $C^{m \times n}$, $Q^{m \times n}$ be the sets of all $m \times n$ real matrices, all $m \times n$ complex matrices, all $m \times n$ quaternion matrices, respectively. For $A \in C^{m \times n}$, $\text{Re}(A)$ and $\text{Im}(A)$ denote the real part and the imaginary part of matrix A , respectively. For $A \in Q^{m \times n}$, A^T , A^H and A^+ denote the transpose matrix, the conjugate transpose matrix and the Moore-Penrose generalized inverse matrix of matrix A , respectively. Let $S_n = (e_n, e_{n-1}, \dots, e_1)$, where e_i represents the i th column of the identity matrix. For any $A \in Q^{m \times n}$ and $\eta \in \{i, j, k\}$, denote $A^{\eta H} = -\eta A^H \eta$. If $A^{\eta H} = A$, then A is called an η -Hermitian matrix [1]. The set of all $n \times n$ quaternion η -Hermitian matrices is denoted by $\eta HQ^{n \times n}$. $\|A\|$ denotes the Frobenius norm of $A \in Q^{m \times n}$. For matrix $A \in Q^{m \times n}$, let $a_i = (a_{1i}, a_{2i}, \dots, a_{ni})$, ($i = 1, 2, \dots, n$), and denote $\text{vec}(A) = (a_1, a_2, \dots, a_n)^T$.

Many results have been achieved about the quaternion matrix equations. See, for example, [2-8] for more details. Matrix equations are important in many applications. For example, the well known Lyapunov equation $AX - XA^H = C$ and Stein equation $X - AXA^H = C$ play important roles in the theory and applications of stability and control. The matrix equation $AXA^H = D$ is very useful in inverse problems in vibration. η -Hermitian matrices is an important class of matrices applied in widely linear modelling due to the quaternion involution properties (see[9-10] for details). In

this paper, using the basic matrices of η -Hermitian matrices, the Moore-Penrose generalized inverse and the complex representation of quaternion matrices, we will consider the following problem.

Problem I. Given $A \in Q^{m \times n}$, $B \in Q^{n \times s}$, $C \in Q^{m \times s}$, let

$$H_L = \left\{ X \mid X \in \eta HQ^{n \times n}, \|AXB - C\| = \min_{\tilde{X} \in \eta HQ^{n \times n}} \|A\tilde{X}B - C\| \right\}. \quad (1)$$

Find $\hat{X} \in H_L$ such that

$$\|\hat{X}\| = \min_{X \in H_L} \|X\|. \quad (2)$$

The solution \hat{X} of Problem I is called the least squares η -Hermitian solution of the quaternion matrix equation $AXB = C$ with the least norm.

2 The Structure of $\eta HQ^{n \times n}$

We begin with the structure of quaternion η -Hermitian matrices.

A quaternion a can be uniquely expressed as $a = a_0 + a_1i + a_2j + a_3k$ with real coefficients a_0, a_1, a_2, a_3 , and $i^2 = j^2 = k^2 = ijk = -1$, and a can be uniquely expressed as $a = c_1 + c_2j$, where c_1 and c_2 are complex numbers. For any $A \in Q^{m \times n}$, A can be uniquely expressed as $A = A_1 + A_2j$, where $A_1, A_2 \in C^{m \times n}$. The complex representation matrix of $A \in Q^{m \times n}$ is denoted by

$$f(A) = \begin{pmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{pmatrix} \in C^{2m \times 2n}, \quad (3)$$

and $f(A)$ is uniquely determined by A . For $A \in Q^{m \times n}$, $B \in Q^{n \times s}$, we have $f(AB) = f(A)f(B)$.

Definition 1. For $A \in Q^{n \times n}$, let $a_1 = (a_{11}, \sqrt{2}a_{21}, \dots, \sqrt{2}a_{n1})$, $a_2 = (a_{22}, \sqrt{2}a_{32}, \dots, \sqrt{2}a_{n2})$, ..., $a_{n-1} = (a_{(n-1)(n-1)}, \sqrt{2}a_{n(n-1)})$, $a_n = a_{nn}$, denote by $vec_S(A)$ the following vector containing the entries of matrix A :

$$vec_S(A) = (a_1, a_2, \dots, a_n)^T \in Q^{\frac{n(n+1)}{2}}. \quad (4)$$

Definition 2. For $B \in Q^{n \times n}$, let $b_1 = (b_{21}, b_{31}, \dots, b_{n1})$, $b_2 = (b_{32}, b_{42}, \dots, b_{n2})$, ..., $b_{n-2} = (b_{(n-1)(n-2)}, b_{n(n-2)})$, $b_{n-1} = b_{n(n-1)}$, denote by $vec_A(B)$ the following vector containing the entries of matrix A :

$$\text{vec}_A(B) = \sqrt{2}(b_1, b_2, \dots, b_{n-2}, b_{n-1})^T \in Q^{\frac{n(n-1)}{2}}. \quad (5)$$

Lemma 1 [11]. Suppose $X \in R^{n \times n}$, then

$$(i) X \in SR^{n \times n} \Leftrightarrow \text{vec}(A) = K_S \text{vec}_S(X), \quad (6)$$

where $\text{vec}_S(X)$ is represented as (4), and

$$K_S =$$

$$\frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}e_1 & e_2 & \cdots & e_{n-1} & e_n & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & e_1 & \cdots & 0 & 0 & \sqrt{2}e_2 & e_3 & \cdots & e_{n-1} & e_n & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & e_2 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & e_1 & 0 & 0 & 0 & \cdots & e_2 & 0 & \cdots & \sqrt{2}e_{n-1} & e_n & 0 \\ 0 & 0 & \cdots & 0 & e_1 & 0 & 0 & & 0 & e_2 & \cdots & 0 & e_{n-1} & \sqrt{2}e_n \end{pmatrix}.$$

$$(ii) X \in ASR^{n \times n} \Leftrightarrow \text{vec}(A) = K_A \text{vec}_A(X), \quad (7)$$

where $\text{vec}_A(X)$ is represented as (5), and

$$K_A = \frac{1}{\sqrt{2}} \begin{pmatrix} e_2 & e_3 & \cdots & e_{n-1} & e_n & 0 & \cdots & 0 & 0 & \cdots & 0 \\ -e_1 & 0 & \cdots & 0 & 0 & e_3 & \cdots & e_{n-1} & e_n & \cdots & 0 \\ 0 & -e_1 & & 0 & 0 & -e_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -e_1 & 0 & 0 & \cdots & -e_2 & 0 & \cdots & e_n \\ 0 & 0 & \cdots & 0 & -e_1 & 0 & \cdots & 0 & -e_2 & \cdots & -e_{n-1} \end{pmatrix}.$$

Obviously, $K_S^T K_S = I_{\frac{n(n+1)}{2}}$, $K_A^T K_A = I_{\frac{n(n-1)}{2}}$.

For $A = A_1 + A_2 j \in Q^{n \times n}$, we denote an identification by the symbol \cong , that is, $A \cong \tilde{A} = (A_1, A_2)$, $\text{vec}(A_1) + \text{vec}(A_2)j = \text{vec}(A) \cong \text{vec}(\tilde{A}) = \begin{pmatrix} \text{vec}(A_1) \\ \text{vec}(A_2) \end{pmatrix}$,

$$\|\text{vec}(A)\| = \|\text{vec}(\tilde{A})\| = \left\| \begin{pmatrix} \text{vec}(A_1) \\ \text{vec}(A_2) \end{pmatrix} \right\|.$$

We denote $\vec{A} = (\text{Re}(A_1), \text{Im}(A_1), \text{Re}(A_2), \text{Im}(A_2))$,

$$\text{vec}(\vec{A}) = \begin{pmatrix} \text{vec}(\text{Re}(A_1)) \\ \text{vec}(\text{Im}(A_1)) \\ \text{vec}(\text{Re}(A_2)) \\ \text{vec}(\text{Im}(A_2)) \end{pmatrix}.$$

Notice that $\| \text{vec}(\tilde{A}) \| = \| \text{vec}(\bar{A}) \|$.

For $X = X_1 + X_2 j \in \eta HQ^{n \times n}$, by Definition 1 and [12], we have

$$\begin{aligned} X^H \eta &= \eta X \\ \Leftrightarrow [(\text{Re } X_1)^T - (\text{Im } X_1)^T i - (\text{Re } X_2)^T j - (\text{Im } X_2)^T k] \eta \\ &= \eta [(\text{Re } X_1) + (\text{Im } X_1)i + (\text{Re } X_2)j + (\text{Im } X_2)k] \\ \Leftrightarrow (\text{Re } X_1)^T &= \text{Re } X_1, \quad (\text{Im } X_1)^T = \begin{cases} -\text{Im } X_1, & \eta = i \\ \text{Im } X_1, & \eta \neq i \end{cases}, \\ (\text{Re } X_2)^T &= \begin{cases} -\text{Re } X_2, & \eta = j \\ \text{Re } X_2, & \eta \neq j \end{cases}, \quad (\text{Im } X_2)^T = \begin{cases} -\text{Im } X_2, & \eta = k \\ \text{Im } X_2, & \eta \neq k \end{cases}. \end{aligned}$$

Lemma 2 [12]. Suppose $X \in Q^{n \times n}$, then

$$X \in \eta HQ^{n \times n} \Leftrightarrow \text{vec}(\bar{X}) = K_{\eta H} \text{vec}_{\eta H}(\bar{X}), \quad (8)$$

where

$$\begin{aligned} K_{iH} &= \begin{pmatrix} K_S & 0 & 0 & 0 \\ 0 & K_A & 0 & 0 \\ 0 & 0 & K_S & 0 \\ 0 & 0 & 0 & K_S \end{pmatrix}, \quad \text{vec}_{iH}(\bar{X}) = \begin{pmatrix} \text{vec}_S(\text{Re } A_1) \\ \text{vec}_A(\text{Im } A_1) \\ \text{vec}_S(\text{Re } A_2) \\ \text{vec}_S(\text{Im } A_2) \end{pmatrix}, \\ K_{jH} &= \begin{pmatrix} K_S & 0 & 0 & 0 \\ 0 & K_S & 0 & 0 \\ 0 & 0 & K_A & 0 \\ 0 & 0 & 0 & K_S \end{pmatrix}, \quad \text{vec}_{jH}(\bar{X}) = \begin{pmatrix} \text{vec}_S(\text{Re } A_1) \\ \text{vec}_S(\text{Im } A_1) \\ \text{vec}_A(\text{Re } A_2) \\ \text{vec}_S(\text{Im } A_2) \end{pmatrix}, \\ K_{kH} &= \begin{pmatrix} K_S & 0 & 0 & 0 \\ 0 & K_S & 0 & 0 \\ 0 & 0 & K_S & 0 \\ 0 & 0 & 0 & K_A \end{pmatrix}, \quad \text{vec}_{kH}(\bar{X}) = \begin{pmatrix} \text{vec}_S(\text{Re } A_1) \\ \text{vec}_S(\text{Im } A_1) \\ \text{vec}_S(\text{Re } A_2) \\ \text{vec}_A(\text{Im } A_2) \end{pmatrix}, \end{aligned}$$

Lemma 3 [11]. If $A = A_1 + A_2 j \in Q^{m \times n}$, $B = B_1 + B_2 j \in Q^{n \times s}$, and $C = C_1 + C_2 j \in Q^{s \times t}$, then

$$\text{vec}(\widetilde{ABC}) = (f(C)^T \otimes A_1, f(Cj)^H \otimes A_2) \begin{pmatrix} \text{vec}(\bar{B}) \\ \text{vec}(-\bar{jBj}) \end{pmatrix}. \quad (9)$$

Lemma 4 [12]. If $A = A_1 + A_2 j \in Q^{m \times n}$, $X = X_1 + X_2 \in \eta HQ^{n \times n}$, $B = B_1 + B_2 \in Q^{n \times s}$, $K_{\eta H}$ and the vector $\text{vec}_{\eta H}(\bar{X})$ are defined in the form (8), and let

$$W = \begin{pmatrix} I_{n^2} & iI_{n^2} & 0 & 0 \\ 0 & 0 & I_{n^2} & iI_{n^2} \\ I_{n^2} & -iI_{n^2} & 0 & 0 \\ 0 & 0 & I_{n^2} & -iI_{n^2} \end{pmatrix}.$$

Then

$$\text{vec}(\widetilde{AXB}) = (f(B)^T \otimes A_1, f(Bj)^H \otimes A_2) W \text{vec}_{\eta H}(\vec{X}). \quad (10)$$

3 The Solution of Problem I

Based on our earlier discussions, we now turn our attention to Problem I. The following notations are necessary for deriving the solutions of Problem I.

For $A = A_1 + A_2 j \in Q^{m \times n}$, $B = B_1 + B_2 j \in Q^{n \times s}$, $C = C_1 + C_2 j \in Q^{m \times s}$, set $Q_1 = (f(B)^T \otimes A_1, f(Bj)^H \otimes A_2)$ and

$$Q = Q_1 W, \quad P_1 = \text{Re}(Q), P_2 = \text{Im}(Q), \quad e = \begin{pmatrix} \text{vec}(\text{Re}(\tilde{C})) \\ \text{vec}(\text{Im}(\tilde{C})) \end{pmatrix}. \quad (11)$$

Let

$$\begin{aligned} R &= (I - P_1^+ P_1) P_2^T, & H &= R^+ + (I - R^+ R) Z P_2 P_1^+ (P_1^+)^T (I - P_2^T R^+), \\ Z &= [I + (I - R^+ R) P_2 P_1^+ (P_1^+)^T P_2^T (I - R^+ R)]^{-1}, \\ S_{11} &= I - P_1^+ P_1 + (P_1^+)^T P_2^T Z (I - R^+ R) P_2 P_1^+, & S_{12} &= -(P_1^+)^T P_2^T (I - R^+ R) Z, \\ S_{22} &= (I - R^+ R) Z. \end{aligned}$$

From the results in [13], we have

$$(P_1, P_2)^+ = \begin{pmatrix} P_1^+ - P_1^+ P_2 H \\ H \end{pmatrix}, \quad I - (P_1, P_2)^+ (P_1, P_2) = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix}.$$

Theorem 5. Let $A = A_1 + A_2 j \in Q^{m \times n}$, $B \in Q^{n \times s}$, $C \in Q^{m \times s}$. Then the set S_L of Problem I can be expressed as

$$H_L = \left\{ X \mid \text{vec}(\vec{X}) = K_{\eta H} (P_1^+ - H^T P_2 P_1^+, H^T) e + K_{\eta H} (I - P_1^+ P_1 - RR^+) y \right\}, \quad (12)$$

where y is an arbitrary vector of appropriate order.

Proof. By Lemma 4, $\|AXB - C\| = \|\text{vec}(\widetilde{AXB}) - \text{vec}(\tilde{C})\| = \left\| \begin{pmatrix} P_1 \\ P_2 \end{pmatrix} \text{vec}_{\eta H}(\vec{X}) - e \right\|$,

we have

$$\text{vec}_{\eta H}(\vec{X}) = (P_1^+ - H^T P_2 P_1^+, H^T) e + (I - P_1^+ P_1 - RR^+) y.$$

The proof is completed.

Theorem 2. Problem I has a unique solution $\widehat{X} \in H_L$, and \widehat{X} can be expressed as

$$\text{vec}(\widehat{X}) = K_{\eta H}(P_1^+ - H^T P_2 P_1^+, H^T) e. \quad (13)$$

Acknowledgments. This work is supported by Guangdong Natural Science Fund of China (No.10452902001005845) and Program for Guangdong Excellent Talents in University, Guangdong Education Ministry, China (LYM10128).

References

1. Took, C.C., Mandic, D.P., Zhang, F.Z.: On the unitary diagonalisation of a special class of quaternion matrices. *Appl. Math. Lett.* 24, 1806–1809 (2011)
2. Huang, L.P.: The matrix equation $AXB-CXD=E$ over the quaternion field. *Linear Algebra Appl.* 234, 197–208 (1996)
3. Jiang, T.S., Wei, M.S.: On a solution of the quaternion matrix equation $X-AXB=C$ and its application. *Acta Math. Sin.* 21, 483–490 (2005)
4. Li, Y.T., Wu, W.J.: Symmetric and skew-antisymmetric solutions to systems of real quaternion matrix equations. *Comput. Math. Applic.* 55, 1142–1147 (2008)
5. Yuan, S.F., Liao, A.P., Lei, Y.: Minimization problem for Hermitian matrices over the quaternion field. *Acta Math. Sci.* 29A, 1298–1306 (2009) (in Chinese)
6. Yuan, S.F., Liao, A.P.: Least squares solution of quaternion matrix equation with the least norm. *Linear Multilinear Algebra* 59(9), 985–998 (2011)
7. Yuan, S.F., Liao, A.P., Duan, X.F.: Least squares tridiagonal Hermitian solution and tridiagonal bi-Hermitian solution of the quaternion matrix equation $AXB=C$ with the least norm. *Numerical Mathematics A Journal of Chinese Universities* 32(4), 353–368 (2010)
8. Yuan, S.F., Liao, A.P., Yao, G.Z.: The matrix nearness problem associated with the quaternion matrix equation $AXA^H + BYB^H = C$. *J. Appl. Math. Comput.* 37, 133–144 (2011)
9. Took, C.C., Mandic, D.P.: Augmented second-order statistics of quaternion random signals. *Signal Processing* 91, 214–224 (2011)
10. Ell, T., Sangwine, S.J.: Quaternion involutions and anti-involutions. *Comput. Math. Applic.* 53, 137–143 (2007)
11. Yuan, S.F., Liao, A.P., Lei, Y.: Least squares Hermitian solution of the matrix equation $(AXB, CXD)=(E, F)$ with the least norm over the skew field of quaternions. *Math. Comput. Modelling* 48, 91–100 (2008)
12. Yuan, S.F., Wang, Q.W.: Two special kinds of least squares solutions for the quaternion matrix equation $AXB+CXD=E$ *Electron. J. Linear Algebra* 23, 257–274 (2012)
13. Magnus, J.R.: L-structured matrices and linear matrix equations. *Linear Multilinear Algebra* 14, 67–88 (1983)

Reasonable Arrangements for Ophthalmic Beds Based on Linear Programming

Lihui Zhou, Shaohong Yan, and Yongsheng Zhang

College of Sciences, Hebei United University, Tangshan, Hebei, China
zhoulh324@163.com

Abstract. In this paper, faced with the status that the waiting queue for the ophthalmic beds is increasingly longer and longer which is caused by FCFS (First come, First serve) rule and the number of new patients per day subject to various types of Poisson distribution, this paper solves the problem of bed arrangements model with the knowledge of the establishment of operations research. Also for a kind of patients occupying beds in a fixed program, the proportional allocation model of our own is established by actual test data and found good effect. Taking the factors of the gradual growth in queue is that the FCFS causes the additional occupied time of the beds into account, in this model, this paper regards the least wasting days as the objective function. Also the constraints, based on the assumption that beds every day are at full basis, are that on the same day ,the patient to be arranged is equal to be discharged and the number of arranged patients is not greater than the accumulative patients. So this paper comes to an arrangement for the least waiting time. According to the arrangement, by simulation, this paper concludes that there are 21 days for wasting of time and 42 times of turnover next week. Compared with 34 times of turnover by FCFS rule, efficiency of arrangement above is improved by 24%.

Keywords: bed turnover times, linear programming, the matrix of wasting time, the matrix of service time.

1 Introduction

Usually, hospitals arrange beds for all non-urgent patients based on the FCFS (First come, First serve) rule, which leads to the queue of waiting patients to be longer and longer. Consequently, hospital hopes that, through mathematical modeling to help solve the issue of reasonable arrangements for inpatient beds to enhance the efficient use of hospital resources.

Take this case into account: The hospital eye clinic is open every day, inpatient beds is 79. Currently the hospital does cataract surgery on Monday and Wednesday, the preparation time before surgery is 1 or 2 days. Compared to the patient with one eye being surge red, the other type is more, accounting for around 60%, for them, one eye is on Monday, and another is on Wednesday.

Trauma patients are usually immediately arranged only if there are free beds, and accept surgeon on the second day. Other diseases are more complex, but generally 2-3

days after hospitalization for surgery and observed for a longer time ever since. These diseases need to arrange the operation time but not normally scheduled on Mondays and Wednesdays. Trauma patients can be ignored when modeling because of its uncommon.

Supposed that the surgical conditions of the hospital is adequate, so you can only consider the arrangement of doctors, usually cataract surgery and other ophthalmic surgery (except Trauma) is not scheduled to do at the same day. For the question of long queue for the hospital issue, this paper will search for a rational evaluation system to evaluate the quality of bed arrangements. And based on the current situation, this paper will establish a reasonable model of bed arrangements, According to the known number of patients; determine which patients should be arranged for hospital the next day.

2 Model Analysis

On the basis of stable inputs of patients and according to the hospital rules, this paper needs to design the hospitalization model of reasonable arrangements.

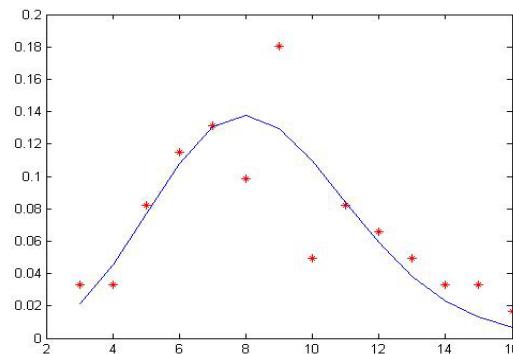
For bed occupancy rate, bed turnover times and the average length of stay are the indicators which reflect the hospital beds used. Thinking of the assumption that beds have been fulfilled and the correlation between bed turnover times and the average length of stay, this article only sees the bed turnover times as the evaluation.

The model needs be considered how to determine the admission arrangement to minimize the waste of that period, because of the gap between the admission time and the best time which is caused by the FCFS rule and the operation time arranged by hospital. Through dealing with the dates in Appendix, this paper gets the time to be discharged for currently 79 patients. Assuming that beds is fulfilled every day, based on the time this paper got just now , this paper can get the number of people discharged from hospital the next period and the best time for patients admitted to the hospital according to the hospital rules. Remember the constraints are that the number of all arranged patients is less than the cumulative number of patients and the number of patients discharged from hospital is equal to those admitted to hospital beds on the same day. Looking on the least waste of time as the objective function, this paper can obtain a reasonable arrangement of beds. Of course, in accordance with the original arrangement of the model, this paper needs to carry out the comparison, evaluation and analysis.

3 Model and Solution for Bed's Arrangements

3.1 The Patient's Number of Each Type Outpatient Clinic Per Day

Count the number of outpatient per day, as shown below, and verify the number obey the Possion distribution:

**Fig. 1.** The number of outpatient per day

The number of patients per day is stable; we can get all kinds of out-patient number per day form its proportion.

Table 1. Out-patient number per day

Retinal diseases	Glaucoma	trauma	Cataract	Cataract(eyes)
3	1	1	2	2

3.2 Make Sure the Best Moment of Hospital

Two days preparation is needed for Retinal diseases, glaucoma, and one day for cataract surgery. The Reasonable arrangements for per kind patient shows in Table 2.

Table 2. Wasted time

	Retinal diseases	Glaucoma	Cataract	Cataract(both eyes)
Mon.	1	1	1	6
Tue.	0	0	0	5
Wed.	0	0	4	4
Thu.	0	0	3	3
Fri.	0	0	2	2
Sat.	1	1	1	1
Sun.	0	0	0	0

3.3 Calculate the Proposed Number of Discharged Patients

Statistic the time from surgery to discharging from hospital for the first time of 349 patients in Appendix.

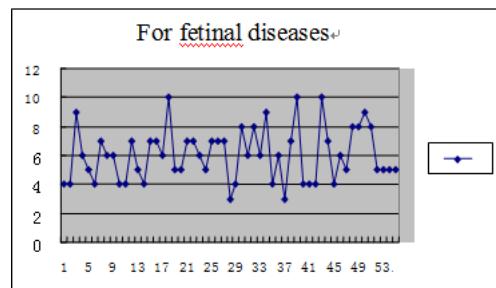


Fig. 2. For fetal diseases

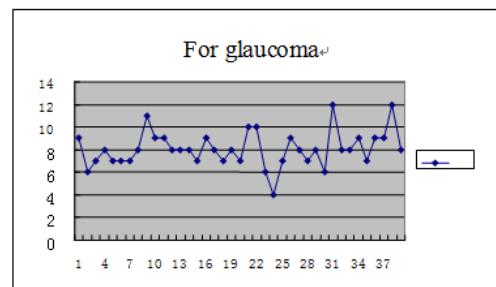


Fig. 3. For glaucoma

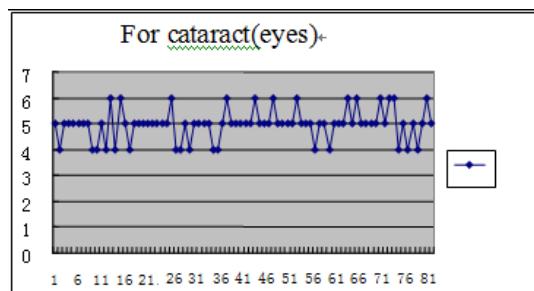


Fig. 4. For cataract (eyes)

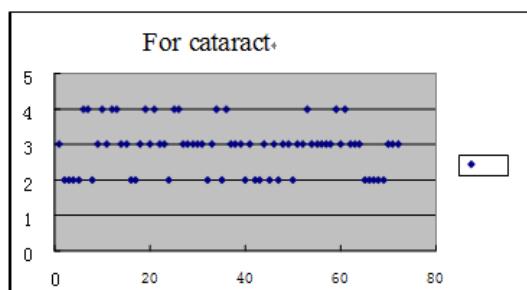


Fig. 5. For cataract

The figure shows a relatively smooth line; take the mean time as the time from surgery to discharge, shown in the following table:

Table 3. The mean time

Retinal diseases	Glaucoma	Trauma	Cataract	Cataract(eyes)
10	8	6	3	5

And then to predict the time and the number of 79 patients discharged from hospital (found that the patients who should discharge from hospital but not are arranged to discharge on the day of 12), shown in the following table:

Table 4. The number of patients discharged from hospital

Discharged	Available beds	Discharged	Available beds
9/12	6	9/18	3
9/13	19	9/19	9
9/14	5	9/20	9
9/15	4	9/21	4
9/16	6	9/22	5
9/17	7	9/23	2

3.4 The Number of Unscheduled for Hospital before September 12

The number of hospitalized patients were not scheduled to obtained total number of waiting, shown in the following table:

Table 5. The total number of waiting

Cataract	Cataract(eyes)	Glaucoma	Retinal diseases	Trauma
21	29	15	36	1

3.5 Optimal Arrangement of Beds

Arrangement of various types of patients within one week less than the cumulative number of patients

$$\sum_{i=1}^7 x_{ij} \leq c_j \times \Delta t + yu(i) \quad (1)$$

The number of patients who are hospitals equal to the number of discharged from hospital: beds are full used.

$$\sum_{j=1}^4 x_{ij} = u(i) \quad (2)$$

Objective function

$$\begin{aligned} \min y &= \sum_{i=1}^7 \sum_{j=1}^4 a_{ij} \times x_{ij} \\ s.t. \sum_{i=1}^7 x_{ij} &\leq c_j \times \Delta t + yu(i) \\ \sum_{j=1}^4 x_{ij} &= u(i) \end{aligned}$$

Where c_j is the number of patient j who went into the clinic per day.

Calculate out the reasonable arrangements' time with Lingo for the first week shown in the following table:

Table 6. The reasonable arrangements' time for the first week

	Retinal diseases	Glaucoma	Cataract	Cataract(both eyes)
	3	0	0	0
Tue.	5	0	0	0
Wed.	6	0	0	0
Thu.	2	0	0	0
Fri.	0	4	0	0
Sat.	0	0	12	6
Sun.	4	0	0	0

In accordance with a model to evaluate the problem $X = 42$, the total available beds' number is 42 for this week.

According to FCFS rule, we can get it working 34, which are efficiency is improved by 24%.

Acknowledgements. The funding organizations are Hebei Higher Social Science Research 2011 Annual Fund(No.SZ2011518) and Tangshan Municipal Bureau of Science and Technology research and development guide plan(No.111302013b), (No.111302012b).

References

1. Zhong, G., Wang, X.: Bed utilization index, objective analysis of the optimal index and rank and the ratio of hospital beds used in evaluating the efficiency of the application. Journal of Hospital Statistics 11, 114–116 (2004)
2. Hu, Y.: Operations Research and its application. Higher Education Press, Beijing (2008)
3. Jiang, Q., Xie, J., Ye, J.: Mathematical model. Higher Education Press, Beijing (2006)
4. Wan, X.: Probability Theory and Mathematical Statistics. Science Press, Beijing (2007)
5. Gao, L., Sun, D., Yue, Z.: Efficiency indicators of utilization of hospital beds hospital beds set up. The Situation of Hospital Statistics (2009)

6. He, X.: Multivariate Statistical Analysis, 2nd edn. Renmin University of China Press, Beijing (2008)
7. Ou, Y., Sheng, Y.: Strategic New Industry Research. Hunan Social Sciences 5, 111–115 (2010)
8. Yang, A., Liu, C., Chang, J., Feng, L.: TOPSIS-Based Numerical Computation Methodology for Intuitionistic Fuzzy Multiple Attribute Decision Making. Nformation—an International Interdisciplinary Journal 14(10), 3169–3174 (2011)
9. Zhou, L., Wang, Y.: The risk warning model of Chinese listed companies Based on Logistic regression and BP neural network. In: International Conference on Computer Application and System Modeling, vol. 2, pp. 607–610 (2010)
10. Zhou, L.: The Principal Component Analysis about Three-dimensional time series data of Chinese information process. In: International Conference on E-Business and E-Government, vol. 2, pp. 4901–4904 (2010)
11. Zhang, C., Liang, R., Liu, L., Wang, J.: Set pair community mining and situation analysis based on web social network. In: International Conference on Advanced in Control Engineering and Information Science, pp. 3456–3460 (2011)

Error Analysis of Generalized LxF Schemes for Linear Advection Equation with Damping

Yun Wu and Dong Qu

College of Science, Jiujiang University
Jiujiang, China, 332005
wuyun19821231@sohu.com

Abstract. Local oscillations existing in the generalized Lax-Friedrichs (LxF) schemes are proposed and analyzed on computing of the linear advection equation with damping. For the discretization of some special initial data under stable conditions, local oscillations in numerical solutions are observed. Three propositions are also raised about how to control those oscillations via some numerical examples. In order to further explain this, discrete Fourier analysis and the modified equation analysis is used to distinguish the dissipative and dispersive effects of numerical schemes for low frequency and high frequency modes, respectively.

Keywords: finite difference schemes, oscillations, discrete Fourier analysis, modified equation analysis.

1 Introduction

Oscillation is an unsurprising phenomenon in numerical computations, but it is a very surprising phenomenon for computing of hyperbolic conservation laws using the Lax-Friedrichs(LxF) scheme. In this thesis, we consider the linear advection equation with damping.

$$u_t + au_x = -\alpha u, \quad x \in R, \alpha > 0. \quad (1)$$

In ([6],[12]), to compute the numerical solution of the hyperbolic conservation laws

$$u_t + f(u)_x = 0, \quad x \in R, t > 0, \quad (2)$$

where $u = (u_1, \dots, u_m)^T$, and $f(u) = (f_1, \dots, f_m)^T$, we consider the generalized LxF scheme of the viscosity form

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}[f(u_{j+1}^n) - f(u_{j-1}^n)] + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (3)$$

where the mesh ratio $\nu = \tau/h$ is assumed to be a constant, τ and h are step sizes in time and space, respectively, u_j^n denotes an approximation of $u(jh, n\tau)$, the term $q \in (0, 1]$ is the coefficient of numerical viscosity. When $q = 1$, it is the classical LxF scheme.

With the flux function $f = au$, 1 is the linear advection equation accordingly

$$u_t + au_x = 0, \quad x \in R, t > 0, \quad (4)$$

and the scheme 3 turns into the generalized Lax-Friedrichs scheme of 4

$$u_j^{n+1} = u_j^n - \frac{a\nu}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n). \quad (5)$$

By adding a damping term $-\alpha u$ (α is a positive constant) to the right of 4, we obtain the linear advection equation with damping 1. We can take account of the discretization of the damping term in the form

$$u_j^{n+1} = u_j^n - \frac{a\nu}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) - \alpha\tau u_j^n \quad (6)$$

2 Analysis of Numerical Dissipation and Phase Error

In this section we attempt to analyze the numerical dissipation and phase error mechanisms of 6. We apply discrete Fourier analysis and the method of modified equation analysis. Both of the methods give complimentary results, which are consistent. We use the method of discrete Fourier analysis (as in [1,2]) to discuss the dissipation and phase error mechanisms of the generalized LxF scheme 6. Denote a Fourier mode using the scaled wave number $\xi = 2\pi kh$ by $e^{i\xi}$. Then using it as initial data for a linear finite difference scheme results after n times steps in the solution

$$u_k^n = \lambda_k^n e^{i\xi} = (\lambda(k))^n e^{i\xi}, \quad i^2 = -1, \quad (7)$$

where λ_k^n is the amplitude. The modulus of the ratio

$$\lambda(k) = \lambda_k^{n+1}/\lambda_k^n$$

is the amplitude of the mode for one time step. For the scheme 6 we have with $\nu = \tau/h$ in particular.

$$\lambda(k) = 1 - \alpha\tau - q(1 - \cos \xi) - ia\nu \sin \xi \quad (8)$$

and

$$\begin{aligned} & |\lambda(k)|^2 \\ &= [1 - \alpha\tau - q(1 - \cos \xi)]^2 + (\nu a)^2 \sin^2 \xi \\ &= (1 - \alpha\tau)^2 + 4[a^2\nu^2 - q(1 - \alpha\tau)] \sin^2(\xi/2) \\ &\quad + 4(q^2 - a^2\nu^2) \sin^4(\xi/2) \end{aligned} \quad (9)$$

Under the condition $0 < \frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq 1 - \alpha\tau$, we have from 8 the estimate

$$\begin{aligned} & |\lambda(k)|^2 \\ &= (1 - \alpha\tau)^2 + 4[a^2\nu^2 - q(1 - \alpha\tau)] (\sin^2(\xi/2) \\ &\quad - \sin^4(\xi/2)) + 4q[q - (1 - \alpha\tau)] \sin^4(\xi/2) \end{aligned} \quad (11)$$

$$\leq 1 \quad (12)$$

Thus these condition imply that the schemes are linearly stable. Conversely, assume that $|\lambda(k)|^2 \leq 1$ then from 8 we have $[a^2\nu^2 - q(1 - \alpha\tau)] + (q^2 - a^2\nu^2)\sin^2(\xi/2) \leq 0$. Now for $\xi = 0$ we obtain $\frac{a^2\nu^2}{1-\alpha\tau} \leq q$. For $\xi = \pi$ this gives $q[q - (1 - \alpha\tau)] \leq 0$ or since we have $0 < \frac{a^2\nu^2}{1-\alpha\tau} \leq q$ we obtain $q \leq 1 - \alpha\tau$. Therefore, GLxF scheme is conditionally stable and $0 < \frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq 1 - \alpha\tau$ is necessary and sufficient for stability.

The exact solution of the Fourier mode $e^{i\xi}$ for $x = h$ after one time step τ is $e^{i(\xi-2\pi ak\tau)} = e^{-i2\pi ak\tau}e^{i\xi} = \lambda_{exact}(k)e^{i\xi}$. We see from 8 that the *amplification error*, i.e. the error in amplitude modulus is of order $O(1)$. Further, comparing the exponents of $\lambda(k)$ and $\lambda_{exact}(k)$ there is a *phase error* $\arg \lambda(k) - (-2\pi ak\tau)$. The *relative phase error* is then defined as

$$E_p(k) := \frac{\arg \lambda(k)}{-2\pi ak\tau} - 1 = -\frac{\arg \lambda(k)}{\nu a \xi} - 1$$

A mode is a low frequency mode if $\xi \approx 0$ and a high frequency mode if $\xi \approx \pi$.

We first look at the low frequency modes

$$(U^s)_j^n := \lambda_k^n e^{i\xi j}, \quad \xi \approx 0$$

Then substituting $(U^s)_j^n$ into 6 yields,

$$\lambda(k) = 1 - \alpha\tau - q(1 - \cos \xi) - i a \nu \sin \xi \quad (13)$$

and

$$\begin{aligned} |\lambda(k)|^2 &= (1 - \alpha\tau - q(1 - \cos \xi))^2 + (\nu a)^2 \sin^2 \xi \\ &= (1 - \alpha\tau)^2 - (1 - \cos \xi)[2q(1 - \alpha\tau) - \\ &\quad q^2(1 - \cos \xi) - a^2\nu^2(1 + \cos \xi)] \end{aligned} \quad (14)$$

We express 14 by using $\cos \xi = 1 - \frac{1}{2}\xi^2 + o(\xi^4) + \dots$

$$|\lambda|^2 = (1 - \alpha\tau)^2 + [a^2\nu^2 - q(1 - \alpha\tau)]\xi^2 + \frac{1}{4}(q^2 - a^2\nu^2)\xi^2 + \dots \quad (15)$$

So, we can see that the dissipation error is of order $O(1)$. From 8 we obtain, as $\frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq a\nu$

$$\frac{d(|\lambda(k)|^2)}{dq} = 2(\cos \xi - 1)[(1 - \alpha\tau) - q(1 - \cos \xi)] < 0 \quad (16)$$

for fixed $\xi \in [0, \pi/2]$. This implies that when $\frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq a\nu$, the dissipation becomes stronger as q larger, and then oscillations become weaker.

Also, we use the method of modified equation analysis to further investigate the mechanisms of dissipation and phase error of (6). This method of modified equation analysis was introduced in [2,6]. We attempt to analyze how the dissipation offsets the oscillations caused by the large phase error.

As in [2], we split the solution $(U)_j^n$ into a smooth part $(U^s)_j^n$ and an oscillatory term $(U^h)_j^n$.

The smooth solution $(U^s)_j^n$ satisfies (6). Here the notation D_{+t} is a forward difference operator in time.

$$\Delta_{+t} = U_j^{n+1} - U_j^n = e^{\tau \partial t} \quad (17)$$

As in [2,Page 170], define $D_{+t} := \Delta_{+t}/\tau$, then we obtain

$$\begin{aligned} \partial_t &= \frac{\ln(1+\Delta_{+t})}{\tau} \\ &= D_{+t} - \frac{1}{2}\tau D_{+t}^2 + \frac{1}{3}\tau^2 D_{+t}^3 - \frac{1}{4}\tau^3 D_{+t}^4 + \dots \end{aligned} \quad (18)$$

In accordance with 6, taking the standard Taylor expansion yields,

$$D_{+t} = \Delta_{+t}/\tau = -\alpha - a\partial_x + \frac{qh^2}{2\tau}\partial_x^2 - \frac{ah^2}{6}\partial_x^3 + \dots \quad (19)$$

Substituting (19) into (18) yields,

$$\begin{aligned} \partial_t &= (-\alpha - a\partial_x + \frac{qh^2}{2\tau}\partial_x^2 - \frac{ah^2}{6}\partial_x^3) \\ &\quad - \frac{\tau}{2}(-\alpha - a\partial_x + \frac{qh^2}{2\tau}\partial_x^2 - \frac{ah^2}{6}\partial_x^3)^2 \\ &\quad + \frac{\tau^2}{3}(-\alpha - a\partial_x + \frac{qh^2}{2\tau}\partial_x^2 - \frac{ah^2}{6}\partial_x^3)^3 + \dots \\ &= \frac{\ln|1-\alpha\tau|}{\tau} - \frac{a}{1-\alpha\tau}\partial_x + \frac{qh^2(1-\alpha\tau)-a^2\tau}{2\tau(1-\alpha\tau)^2}\partial_x^2 \\ &\quad + \frac{ah^2[(1-\alpha\tau)(3q-1+\alpha\tau)-2a^2\nu^2]}{6(1-\alpha\tau)^3}\partial_x^3 + \dots \end{aligned} \quad (20)$$

Here the notation \tilde{U}^s is used to express the solution of exact differential equation corresponding to (1.6), we derive the modified equation about the smooth part:

$$\begin{aligned} \partial_t \tilde{U}^s &+ \frac{a}{1-\alpha\tau}\partial_x \tilde{U}^s \\ &= \frac{\ln|1-\alpha\tau|}{\tau} \tilde{U}^s + \frac{h^2}{2\tau} \frac{q(1-\alpha\tau)-a^2\nu^2}{(1-\alpha\tau)^2} \partial_x^2 \tilde{U}^s \\ &\quad + \frac{ah^2}{6} \frac{(1-\alpha\tau)(3q-1+\alpha\tau)-2a^2\nu^2}{(1-\alpha\tau)^3} \partial_x^3 \tilde{U}^s + \dots \end{aligned} \quad (21)$$

In the modified equation, the second order term $\frac{h^2}{2\tau} \frac{q(1-\alpha\tau)-a^2\nu^2}{(1-\alpha\tau)^2} \partial_x^2 \tilde{U}^s$ is the numerical viscosity. When $0 < \frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq 1 - \alpha\tau$, we have $q(1 - \alpha\tau) - a^2\nu^2 \geq 0$, thus the numerical viscosity becomes stronger as q is larger. The conclusion is the same as Fourier analysis for low frequency modes.

Now we use the modified equation analysis to study the oscillatory part. The oscillatory solution $(U^h)_j^n$ is written as

$$(U^h)_j^n = (-1)^{j+n}(U^o)_j^n,$$

where $(U^o)_j^n$ is viewed as the perturbation amplitude of the chequerboard mode. Then the oscillatory term $(U^o)_j^n$ satisfies ,in accordance with (1.6),

$$\begin{aligned} & (-1)^{j+n+1}(U^o)_j^{n+1} \\ &= (-1)^{j+n}(U^o)_j^n - \frac{a\nu}{2}[(-1)^{j+n+1}(U^o)_{j+1}^n \\ &\quad - (-1)^{j+n-1}(U^o)_{j-1}^n] + \frac{q}{2}[(-1)^{j+n+1}(U^o)_{j+1}^n \\ &\quad - 2(-1)^{j+n}(U^o)_j^n + (-1)^{j+n-1}(U^o)_{j-1}^n] \\ &\quad - \alpha\tau(-1)^{j+n}(U^o)_j^n \end{aligned} \quad (22)$$

that is ,

$$\begin{aligned} & (U^o)_j^{n+1} - (U^o)_j^n \\ &= (q + \alpha\tau - 2)(U^o)_j^n - \frac{a\nu}{2}[(U^o)_{j+1}^n - (U^o)_{j-1}^n] \\ &\quad + \frac{q}{2}[(U^o)_{j+1}^n + (U^o)_{j-1}^n] \end{aligned} \quad (23)$$

We use the notation $\tilde{U}^o(jh, n\tau)$ to express $(U^o)_j^n$ inserted into (23) and apply the standard approach. That is , taking the standard Taylor expansion yields

$$\Delta_{+t}\tilde{U}^o = (2q + \alpha\tau - 2)\tilde{U}^o - a\tau\partial_x\tilde{U}^o + \frac{qh^2}{2}\partial_x^2\tilde{U}^o - \frac{a\nu h^3}{6}\partial_x^3\tilde{U}^o + \dots \quad (24)$$

where $D_{+t} = \frac{\Delta_{+t}}{\tau} = \frac{e^{\tau\partial_t} - 1}{\tau}$. Next we write (24) as

$$(e^{\tau\partial_t} - 1)\tilde{U}^o = \beta\tilde{U}^o - a\tau\partial_x\tilde{U}^o + \frac{qh^2}{2}\partial_x^2\tilde{U}^o - \frac{a\tau h^2}{6}\partial_x^3\tilde{U}^o + \dots$$

where $\beta = 2q + \alpha\tau - 2$. Note the following basic facts, namely the formal operator expansion

$$\tau\partial_t = \ln((e^{\tau\partial_t} - 1) + 1) = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(e^{\tau\partial_t} - 1)^m}{m},$$

and the well known power series

$$\frac{1}{(1+z)^2} = \sum_{m=0}^{\infty} (-1)^m (m+1) z^m, \quad \text{where } z \in (-1, 1),$$

$$\frac{1}{(1+z)^3} = \sum_{m=0}^{\infty} (-1)^m \frac{(m+1)(m+2)}{2} z^m, \quad \text{where } z \in (-1, 1),$$

Let $C_m^l = \frac{m!}{(m-l)!l!}$ denote the binomial coefficients for $l \leq m$. We obtain ,by ignoring terms of orders higher than three, that

$$\begin{aligned}
& \tau \partial_t \tilde{U}^o \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} [(e^{\tau \partial_t} - 1) \tilde{U}^o]^m \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (\beta \tilde{U}^o - a \tau \partial_x \tilde{U}^o \\
&\quad + \frac{q h^2}{2} \partial_x^2 \tilde{U}^o - \frac{a \tau h^2}{6} \partial_x^3 \tilde{U}^o + \dots)^m \\
&= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \beta^m \tilde{U}^o + \\
&\sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 \beta^{m-1} (-a \tau \partial_x) \tilde{U}^o \\
&\quad + \left\{ \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m} C_m^2 \beta^{m-2} a^2 \tau^2 \partial_x^2 \right. \\
&\quad \left. + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 \beta^{m-1} \frac{q h^2}{2} \partial_x^2 \right\} \tilde{U}^o \\
&\quad + \left\{ \sum_{m=3}^{\infty} \frac{(-1)^{m+1}}{m} C_m^3 \beta^{m-3} (-a \tau \partial_x)^3 \right. \\
&\quad \left. + \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 C_{m-1}^1 \beta^{m-2} (-a \tau \partial_x) \frac{q h^2}{2} \partial_x^2 \right. \\
&\quad \left. + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 \beta^{m-1} \left(-\frac{a \tau h^2}{6} \partial_x^3 \right) \right\} \tilde{U}^o + \dots \\
&= \ln |\beta + 1| \tilde{U}^o - \frac{a \tau}{1+\beta} \partial_x \tilde{U}^o \\
&\quad + \left\{ -\frac{a^2 \tau^2}{2(1+\beta)^2} + \frac{q h^2}{2(1+\beta)} \right\} \partial_x^2 \tilde{U}^o \\
&\quad + \left\{ \frac{-a^3 \tau^3}{3(1+\beta)^3} + \frac{a \tau}{(1+\beta)^2} \frac{q h^2}{2} - \frac{1}{1+\beta} \right. \\
&\quad \left. \frac{a \tau h^2}{6} \right\} \partial_x^3 \tilde{U}^o + \dots \\
&= \ln |2q + \alpha \tau - 1| \tilde{U}^o - \frac{a \tau}{2q + \alpha \tau - 1} \partial_x \tilde{U}^o \\
&\quad + \frac{1}{2} \frac{[q h^2 (2q + \alpha \tau - 1) - a^2 \tau^2]}{(2q + \alpha \tau - 1)^2} \partial_x^2 \tilde{U}^o \\
&\quad + \frac{a \tau}{6} \frac{[h^2 (2q + \alpha \tau - 1)(q + 1 - \alpha \tau) - 2a^2 \tau^2]}{(2q + \alpha \tau - 1)^3} \partial_x^3 \tilde{U}^o + \dots \tag{25}
\end{aligned}$$

Thus we derive the modified equation for the oscillatory part:

$$\begin{aligned}
& \partial_t \tilde{U}^o + \frac{a}{2q + \alpha \tau - 1} \partial_x \tilde{U}^o \\
&= \frac{\ln |2q + \alpha \tau - 1|}{\tau} \tilde{U}^o \\
&\quad + \frac{h^2}{2\tau} \frac{[q(2q + \alpha \tau - 1) - a^2 \nu^2]}{(2q + \alpha \tau - 1)^2} \partial_x^2 \tilde{U}^o \\
&\quad + \frac{ah^2}{6} \frac{[(2q + \alpha \tau - 1)(q + 1 - \alpha \tau) - 2a^2 \nu^2]}{(2q + \alpha \tau - 1)^3} \partial_x^3 \tilde{U}^o + \dots \tag{26}
\end{aligned}$$

where $(1 - \alpha \tau)/2 < q < 1 - \alpha \tau$,

We call the zero term in (26) $\frac{\ln |2q + \alpha \tau - 1|}{\tau} \tilde{U}^o$ a numerical damping term and the second order term $\frac{h^2}{2\tau} \frac{[q(2q + \alpha \tau - 1) - a^2 \nu^2]}{(2q + \alpha \tau - 1)^2} \partial_x^2 \tilde{U}^o$ a numerical viscosity. They play

distinct dissipation roles in controlling the amplitude of high frequency modes. But this dissipation is not enough to control numerical oscillations caused by the relative phase errors of high frequency modes, and only the numerical damping can suppress the oscillation.

With the stability condition of the scheme 6, we obtain $(1-\alpha\tau)/2 < q < 1-\alpha\tau$, several conclusions are in order.

(1) As $(1-\alpha\tau)/2 < q < 1-\alpha\tau$, there is a strong damping term $\frac{\ln|2q+\alpha\tau-1|}{\tau}$ ($\ln(2q+\alpha\tau-1) < 0$) to suppress the oscillation. The numerical damping becomes stronger as q decrease.

(2) In particular , if $q = \frac{1-\alpha\tau}{2}$, the oscillation is damped out immediately, by noting that

$$\lim_{q \rightarrow (1-\alpha\tau)/2+0} \frac{\ln|2q+\alpha\tau-1|}{\tau} = -\infty$$

So the numerical damping becomes infinite for $q = \frac{1-\alpha\tau}{2}$.

3 Discussion and Conclusion

In the present paper, we are discussing the local oscillations in the particular generalized LxF scheme. We have individually analyzed the resolution of the low and high frequency modes $u_j^n = \lambda_k^n e^{ijk\xi}$, $\xi = 2\pi kh$ in numerical solutions. Our approach is the discrete Fourier analysis and the modified equation analysis, which are applied to investigating the numerical dissipative and dispersive mechanisms as well as relative phase errors.

1. Relative phase error. For the low frequency modes, the error is of order $O(1)$, while for high frequency modes the error is of order $O(1)$ after each time step, which is generally independent of the parameter q .

2. Numerical dissipation. For the low frequency modes, the dissipation is usually of order $O(1)$ for the scheme (1.6), which closely depends on the parameter q . For high frequency modes, the scheme usually has the numerical damping of order $O(1)$ that becomes stronger as q is closer to $\frac{1-\alpha\tau}{2}$, unless it vanishes for the limit case ($q = 1 - \frac{\alpha\tau}{2}$).

Thus we obtain that the relative phase errors should be at least offset by the numerical dissipation of the same order. Otherwise the oscillation could be caused.

We also get the following conclusions.

1. The GLxF scheme 6 is conditionally stable and $0 < \frac{a^2\nu^2}{1-\alpha\tau} \leq q \leq 1 - \alpha\tau$ is necessary and sufficient for stability.

2. Under the stable condition, the oscillation is connected with these factors: If $q > \frac{1-\alpha\tau}{2}$, the oscillation becomes weaker as q decrease. If $q > \frac{1-\alpha\tau}{2}$, the oscillation becomes weaker as q increase; The oscillation becomes weaker as α decrease; The oscillation becomes weaker as h decrease; If the initial data can be discretized as square signal, which be discretized with an odd number of grid points, the chequerboard mode (i.e. the oscillation)is present. In contrast, the chequerboard mode (i.e. the oscillation)is suppressed.

3. Once adding the damping into linear advection equation, it is clear from the figures that the damping has resulted in a slight reduction of the modes' height; We also can find even large damping, the oscillation becomes weaker as time goes by, that is to say the chequerboard mode decay.

4. When $q = \frac{1-\alpha\tau}{2}$, the oscillation vanish. When $q = 1 - \frac{\alpha\tau}{2}$, the oscillation is the strongest.

Acknowledgment. The authors would like to thank Prof. Jiequan Li for his helpful remarks and suggestions.

References

1. Li, J., Yang, Z., Zheng, Y.: Characteristic decompositions and interactions of rarefaction waves of 2-D Euler Equations. *Journal of Differential Equations* 250, 782–798 (2011)
2. Li, J., Li, Q., Xu, K.: Comparison of the Generalized Riemann Solver and the Gas-Kinetic scheme for Compressible Inviscid Flow Simulations. *Journal of Computational Physics* 230, 5080–5099 (2011)
3. Dou, L., Dou, J.: The Grid: Time-domain analysis of lossy multiconductor transmission lines based on the LaxCwendroff technique. *Analog Integrated Circuits and Signal Processing* 68, 85–92 (2011)
4. Li, J., Liu, T., Sun, Z.: Implementation of the GRP scheme for computing spherical compressible fluid flows. *Journal of Computational Physics* 228, 5867–5887 (2009)
5. Zhu, P., Zhou, S.: Relaxation Lax-Friedrichs sweeping scheme for static Hamilton-Jacobi equations. *Numer. Algor.* 54, 325–342 (2010)
6. Li, J.-Q., Tang, H.-Z., Warnecke, G., Zhang, L.-M.: Local Oscillations in Finite Difference Solutions of Hyperbolic Conservation Laws. *Mathematics of Computation*, S0025-5718(09)02219-4
7. Morton, K.W., Mayers, D.F.: Numerical Solution of Partial Differential Equations, 2nd edn. Cambridge University Press (2005)
8. Breuss, M.: The correct use of the Lax-Friedrichs method. *M2AN Math. Model. Numer. Anal.* 38, 519–540 (2004)
9. Breuss, M.: An analysis of the influence of data extrema on some first and second order central approximations of hyperbolic conservation laws. *M2AN Math. Model. Numer. Anal.* 39, 965–993 (2005)
10. Thomas, J.W.: Numerical Partial Differential Equations: Finite Difference Methods. Springer (1995)
11. Warming, R.F., Hyett, B.J.: The modified equation approach to the stability and accuracy of finite difference methods. *J. Comput. Phys.* 14, 159–179 (1974)
12. Tadmor, E.: Numerical viscosity and the entropy condition for conservative difference schemes. *Math. Comp.* 43, 369–381 (1984)

Generalized Lax-Friedrichs Scheme for Convective-Diffusion Equation*

Wei Tong and Yun Wu

College of Science, Jiujiang University
Jiujiang, China, 332005
wuyun19821231@sohu.com

Abstract. A new type of generalized Lax-Friedrichs scheme for the convective-diffusion equation $u_t + au_x = \varepsilon u_{xx}$ ($a \in R, \varepsilon > 0$) is given and analyzed. For the convection term, the scheme uses generalized Lax-Friedrichs scheme. For the diffusion term, it uses implicit central difference scheme. The scheme is discussed by applying modified equation analysis, in order to find the relative phase error, numerical dissipation, numerical viscosity, numerical damping and oscillations.

Keywords: generalized Lax-Friedrichs (LxF) scheme, modified equation analysis, oscillations, numerical dissipation, numerical damping, numerical viscosity.

1 Introduction

In [1], to compute the numerical solution of the hyperbolic conservation laws

$$u_t + f(u)_x = 0, x \in R, t > 0, \quad (1)$$

where $u = (u_1, \dots, u_m)^T$, and $f(u) = (f_1, \dots, f_m)^T$, we consider the generalized Lax-Friedrichs(LxF) scheme of the viscosity form

$$u_j^{n+1} = u_j^n - \frac{\nu}{2}[f(u_{j+1}^n) - f(u_{j-1}^n)] + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (2)$$

where the mesh ratio $\nu = \tau/h$ is assumed to be a constant, τ and h are step sizes in time and space, respectively, u_j^n denotes an approximation of $u(jh, n\tau)$, the term $q \in (0, 1]$ is the coefficient of numerical viscosity. When $q = 1$, it is the classical Lax-Friedrichs(LxF) scheme.

With the flux function $f = au$, (1) is the linear advection equation as follows

$$u_t + au_x = 0, \quad x \in R, t > 0, \quad (3)$$

and the scheme (2) turns into the generalized LxF scheme of equation (3)

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n). \quad (4)$$

* Supported by Natural Science Foundation of Jiujiang University.

By adding a diffusion term εu_{xx} (ε is a positive constant) to the right of (3), we obtain a convective-diffusion equation.

$$u_t + au_x = \varepsilon u_{xx}, \quad a \in R, \varepsilon > 0. \quad (5)$$

There are two different finite difference schemes of the convective-diffusion equation (5). For the convective term, we still use the generalized LxF scheme. Then, we have two following ways to approximate the diffusion term: one uses explicit central difference scheme, i.e.

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) + \frac{\varepsilon \tau}{h^2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (6)$$

and the other one uses implicit central difference scheme, i.e.

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) + \frac{\varepsilon \tau}{h^2}(u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}). \quad (7)$$

Scheme (6) also can be written in this form

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \left(\frac{q}{2} + \mu\right)(u_{j+1}^n - 2u_j^n + u_{j-1}^n), \quad (8)$$

where $\mu = \frac{\varepsilon \tau}{h^2}$. Similarly, scheme (7) can be written like this

$$u_j^{n+1} = u_j^n - \frac{\nu a}{2}(u_{j+1}^n - u_{j-1}^n) + \frac{q}{2}(u_{j+1}^n - 2u_j^n + u_{j-1}^n) + \mu(u_{j+1}^{n+1} - 2u_j^{n+1} + u_{j-1}^{n+1}). \quad (9)$$

Scheme (8) and scheme (9) are both the generalized LxF schemes of the convective-diffusion equation(1.5). The term q and ε (in term μ) are called the coefficient of numerical viscosity and the coefficient of physical viscosity, respectively.

2 Chequerboard Modes in the Initial Discretization

As observed in [1], we discussed the discretization of initial data

$$u(x, 0) = u_0(x), \quad x \in [0, 1], \quad (10)$$

with M grid points and $h = 1/M$, while M is even, and $u_0(0) = u_0(1)$. The numerical solution value at the grid point x_j is denoted by u_j^0 . We express this grid point value u_j^0 by using the usual discrete Fourier sums, as in [5, Page 120], and obtain

$$u_j^0 = \sum_{k=-M/2+1}^{M/2} c_k^0 e^{i\xi j}, \quad i^2 = -1, \quad j = 0, 1, \dots, M-1, \quad (11)$$

where $\xi = 2\pi kh$. And the coefficients c_k^0 are expressed as

$$c_k^0 = \frac{1}{M} \sum_{k=0}^{M-1} u_j^0 e^{-i\xi j}, \quad k = -M/2 + 1, \dots, M/2, \quad (12)$$

we have two different types of discretizations, if they are discretized with an odd number of grid points, the chequerboard mode is present. In contrast, if they are discretized with an even number of grid points, the chequerboard mode is suppressed.

(i) Discretization with an odd number of grid points:

Take $j_1, j_2 \in N$ such that $j_1 + j_2$ is an even number. We set $x^{(1)} = (\frac{M}{2} - j_1)h$ and $x^{(2)} = (\frac{M}{2} + j_2)h$. We discretize the square signal (1.12) with $p := j_1 + j_2 + 1$ nodes, such that

$$u_j^0 = \begin{cases} 1, & \text{if } j = M/2 - j_1, \dots, M/2 + j_2, \\ 0, & \text{otherwise.} \end{cases} \quad (13)$$

In [1], we have

$$u_j^0 = (-1)^{j+j_1+M/2}h + ph + \sum_{k \neq 0, M/2} \frac{(-1)^k e^{i\xi(j+j_1)}(1 - e^{-i\xi p})}{M(1 - e^{-i\xi})} \quad (14)$$

(ii) Discretization with an even number of grid points:

Similarly, we use $p - 1 = j_1 + j_2$ even number of grid points to express the square signal in (13) as follows

$$u_j^0 = \begin{cases} 1, & \text{if } j = M/2 - j_1 + 1, \dots, M/2 + j_2, \\ 0, & \text{otherwise.} \end{cases} \quad (15)$$

Then the initial data (15) can be written as follow in [1]

$$u_j^0 = 0 \times (-1)^j + (p - 1)h + \sum_{k \neq 0, M/2} \frac{(-1)^k e^{i\xi(j+j_1-1)}[1 - e^{-i\xi(p-1)}]}{M(1 - e^{-i\xi})}. \quad (16)$$

3 Modified Equation Analysis

As the amplitude error and relative phase error of the Fourier modes have a correspondence with dissipation and phase error mechanisms displayed by related partial differential equations, we use the method of modified equation analysis to further investigate the mechanisms of dissipation and phase error of scheme (1.9). The same to the modified equation analysis for scheme(8), we want to see how the dissipation offsets the large phase error of high frequency modes.

We use implicit central difference scheme to difference the diffusion term εu_{xx} in scheme (1.9), which will give rise to further problem in the modified equation analysis as follows. Similarly, we consider a smooth solution $(U^s)_j^n$ and an oscillatory solution $(U^h)_j^n = (-1)^{j+n}(U^o)_j^n$ respectively, where

$$U_j^n = (U^s)_j^n + (-1)^{j+n}(U^o)_j^n$$

.

3.1 Low Frequency Modes

The smooth solution $(U^s)_n^j$ satisfies scheme (1.9). Implying the same labels $D_{+t} := \Delta_{+t}/\tau$, $\Delta_{+t} = e^{\tau\partial t} - 1$, then we have (17), i.e.

$$\partial_t = D_{+t} - \frac{1}{2}\tau D_{+t}^2 + \frac{1}{3}\tau^2 D_{+t}^3 - \frac{1}{4}\tau^3 D_{+t}^4 + \dots \quad (17)$$

In accordance with (9),

$$\begin{aligned} (U^s)_j^{n+1} &= (U^s)_j^n - \frac{\nu a}{2}[(U^s)_{j+1}^n - (U^s)_{j-1}^n] + \frac{q}{2}[(U^s)_{j+1}^n - 2(U^s)_j^n + (U^s)_{j-1}^n] \\ &\quad + \mu[(U^s)_{j+1}^{n+1} - 2(U^s)_j^{n+1} + (U^s)_{j-1}^{n+1}] \end{aligned} \quad (18)$$

Similar to $\Delta_{+t} = e^{\tau\partial t} - 1$, $\Delta_{+x} = e^{h\partial x} - 1$ is used as follows,

$$\begin{aligned} (U^s)_j^{n+1} &= (U^s)_j^n + [-\frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + \frac{q}{2}(e^{h\partial x} + e^{-h\partial x} - 2)](U^s)_j^n \\ &\quad + \mu(e^{h\partial x} + e^{-h\partial x} - 2)(U^s)_j^{n+1}, \end{aligned}$$

$$\begin{aligned} &[1 - \mu(e^{h\partial x} + e^{-h\partial x} - 2)](U^s)_j^{n+1} \\ &= [1 - \frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + \frac{q}{2}(e^{h\partial x} + e^{-h\partial x} - 2)](U^s)_j^n, \end{aligned}$$

$$\begin{aligned} &[1 - \mu(e^{h\partial x} + e^{-h\partial x} - 2)][(U^s)_j^{n+1} - (U^s)_j^n] \\ &= [-\frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + (\frac{q}{2} + \mu)(e^{h\partial x} + e^{-h\partial x} - 2)](U^s)_j^n, \\ &[1 - \mu(e^{h\partial x} + e^{-h\partial x} - 2)]\Delta_{+t}(U^s)_j^n \\ &= [-\frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + (\frac{q}{2} + \mu)(e^{h\partial x} + e^{-h\partial x} - 2)](U^s)_j^n. \end{aligned} \quad (19)$$

From (19), taking the standard Taylor expansion yields, we obtain

$$\begin{aligned} \Delta_{+t} &= \frac{-\frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + (\frac{q}{2} + \mu)(e^{h\partial x} + e^{-h\partial x} - 2)}{1 - \mu(e^{h\partial x} + e^{-h\partial x} - 2)} \\ &= \frac{-\frac{\nu a}{2}(2h\partial_x + \frac{2h^3}{3!}\partial_x^3 + \dots) + (\frac{q}{2} + \mu)(\frac{2h^2}{2!}\partial_x^2 + \frac{2h^4}{4!}\partial_x^4 + \dots)}{1 - \mu(\frac{2h^2}{2!}\partial_x^2 + \frac{2h^4}{4!}\partial_x^4 + \dots)} \\ &= \frac{-\nu ah\partial_x + (\frac{q}{2} + \mu)h^2\partial_x^2 - \frac{\nu a}{6}h^3\partial_x^3 + \frac{q+2\mu}{24}h^4\partial_x^4 + \dots}{1 - \mu h^2\partial_x^2 - \frac{\mu}{12}h^4\partial_x^4 - \dots} \\ &= -\nu ah\partial_x + (\frac{q}{2} + \mu)h^2\partial_x^2 - (\frac{1}{6} + \mu)\nu ah^3\partial_x^3 + \dots, \end{aligned} \quad (20)$$

then,

$$D_{+t} = \frac{\Delta_{+t}}{\tau} = \frac{1}{\tau}[-\nu ah\partial_x + (\frac{q}{2} + \mu)h^2\partial_x^2 - (\frac{1}{6} + \mu)\nu ah^3\partial_x^3 + \dots]. \quad (21)$$

Substituting (21) into (20), we obtain

$$\partial_t = -a\partial_x + [\varepsilon + \frac{h^2}{2\tau}(q - \nu^2 a^2)]\partial_x^2 + \frac{ah^2}{6}(3q - 2\nu^2 a^2 - 1)\partial_x^3 + \dots, \quad (22)$$

where $\mu = \frac{\varepsilon\tau}{h^2}$, $\nu = \tau/h$.

From (22), we derive the modified equation about the smooth part, where \tilde{U}^s is used to express the solution of exact differential equation corresponding to (9),

$$\begin{aligned} \partial_t \tilde{U}^s + a\partial_x \tilde{U}^s - \varepsilon \partial_x^2 \tilde{U}^s &= \frac{h^2}{2\tau}(q - \nu^2 a^2)\partial_x^2 \tilde{U}^s \\ &\quad + \frac{ah^2}{6}(3q - 2a^2\nu^2 - 1)\partial_x^3 \tilde{U}^s + \dots \end{aligned} \quad (23)$$

The second order term $\frac{h^2}{2\tau}(q - \nu^2 a^2)\partial_x^2 \tilde{U}^s$ in (23) represents the numerical viscosity of (9). Under the condition $q > \nu^2 a^2$, the dissipation effect of the scheme (8) becomes stronger as q increases, which is observed by the Fourier analysis.

3.2 High Frequency Modes

The oscillatory solution $(U^h)_j^n$ is written as

$$(U^h)_j^n = (-1)^{j+n}(U^o)_j^n,$$

where $(U^o)_j^n$ is viewed as the perturbation amplitude of the chequerboard mode. Then the oscillatory term $(U^o)_j^n$ satisfies, in accordance with (1.9),

$$\begin{aligned} &(-1)^{j+n+1}(U^o)_j^{n+1} \\ &= (-1)^{j+n}(U^o)_j^n - \frac{\nu a}{2}[(-1)^{j+n+1}(U^o)_{j+1}^n - (-1)^{j+n-1}(U^o)_{j-1}^n] \\ &\quad + \frac{q}{2}[(-1)^{j+n+1}(U^o)_{j+1}^n - 2(-1)^{j+n}(U^o)_j^n + (-1)^{j+n-1}(U^o)_{j-1}^n] \\ &\quad + \mu[(-1)^{j+n+2}(U^o)_{j+1}^{n+1} - 2(-1)^{j+n+1}(U^o)_j^{n+1} + (-1)^{j+n}(U^o)_{j-1}^{n+1}], \end{aligned}$$

that is ,

$$\begin{aligned} (U^o)_j^{n+1} &= -(U^o)_j^n - \frac{\nu a}{2}[(U^o)_{j+1}^n - (U^o)_{j-1}^n] + \frac{q}{2}[(U^o)_{j+1}^n + 2(U^o)_j^n + (U^o)_{j-1}^n] \\ &\quad - \mu[(U^o)_{j+1}^{n+1} + 2(U^o)_j^{n+1} + (U^o)_{j-1}^{n+1}]. \end{aligned}$$

Now using the same methods about low frequency modes, we deduce as follows,

$$\begin{aligned} &(1 + 4\mu)(U^o)_j^{n+1} + \mu[(U^o)_{j+1}^{n+1} - 2(U^o)_j^{n+1} + (U^o)_{j-1}^{n+1}] \\ &= (2q - 1)(U^o)_j^n - \frac{\nu a}{2}[(U^o)_{j+1}^n - (U^o)_{j-1}^n] + \frac{q}{2}[(U^o)_{j+1}^n - 2(U^o)_j^n + (U^o)_{j-1}^n], \end{aligned}$$

$$\begin{aligned} &[1 + 4\mu + \mu(e^{h\partial x} + e^{-h\partial x} - 2)](U^o)_j^{n+1} \\ &= [2q - 1 - \frac{\nu a}{2}(e^{h\partial x} - e^{-h\partial x}) + \frac{q}{2}(e^{h\partial x} + e^{-h\partial x} - 2)](U^o)_j^n, \end{aligned}$$

$$\begin{aligned} & [1 + 4\mu + \mu(e^{h\partial_x} + e^{-h\partial_x} - 2)]((U^o)_j^{n+1} - (U^o)_j^n) \\ &= [2q - 4\mu - 2 - \frac{\nu a}{2}(e^{h\partial_x} - e^{-h\partial_x}) + (\frac{q}{2} - \mu)(e^{h\partial_x} + e^{-h\partial_x} - 2)](U^o)_j^n, \end{aligned}$$

$$\begin{aligned} & [1 + 4\mu + \mu(e^{h\partial_x} + e^{-h\partial_x} - 2)]\Delta_{+t}(U^o)_j^n \\ &= [2q - 4\mu - 2 - \frac{\nu a}{2}(e^{h\partial_x} - e^{-h\partial_x}) + (\frac{q}{2} - \mu)(e^{h\partial_x} + e^{-h\partial_x} - 2)](U^o)_j^n. \end{aligned} \quad (24)$$

From (24), taking the standard Taylor expansion yields, we have

$$\begin{aligned} \Delta_{+t} &= \frac{2q - 4\mu - 2 - \frac{\nu a}{2}(e^{h\partial_x} - e^{-h\partial_x}) + (\frac{q}{2} - \mu)(e^{h\partial_x} + e^{-h\partial_x} - 2)}{1 + 4\mu + \mu(e^{h\partial_x} + e^{-h\partial_x} - 2)} \\ &= \frac{2q - 4\mu - 2 - \nu ah\partial_x + \frac{q-2\mu}{2}h^2\partial_x^2 - \frac{\nu a}{6}h^3\partial_x^3 + \frac{q-2\mu}{24}h^4\partial_x^4 + \dots}{1 + 4\mu + \mu h^2\partial_x^2 + \frac{\mu}{12}h^4\partial_x^4 - \dots} \\ &= \frac{2q - 4\mu - 2}{1 + 4\mu} - \frac{\nu ah}{1 + 4\mu}\partial_x + \frac{q + 2\mu}{2(1 + 4\mu)^2}h^2\partial_x^2 - \frac{1 - 2\mu}{6(1 + 4\mu)^2}\nu ah^3\partial_x^3 + \dots \end{aligned} \quad (25)$$

We write (25) as

$$(e^{\tau\partial_t} - 1)\tilde{U}^o = P - H\partial_x + Q\partial_x^2 - R\partial_x^3 + \dots,$$

where $P = \frac{2q-4\mu-2}{1+4\mu}$, $H = \frac{\nu ah}{1+4\mu}$, $Q = \frac{q+2\mu}{2(1+4\mu)^2}h^2$, $R = \frac{1-2\mu}{6(1+4\mu)^2}\nu ah^3$. According to the following basic facts and the formal operator expansion

$$\tau\partial_t = \ln((e^{\tau\partial_t} - 1) + 1) = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(e^{\tau\partial_t} - 1)^m}{m},$$

and the well known power series

$$\frac{1}{(1+z)^2} = \sum_{m=0}^{\infty} (-1)^m(m+1)z^m, \quad \text{where } z \in (-1, 1),$$

$$\frac{1}{(1+z)^3} = \sum_{m=0}^{\infty} (-1)^m \frac{(m+1)(m+2)}{2} z^m, \quad \text{where } z \in (-1, 1),$$

Let $C_m^l = \frac{m!}{(m-l)!l!}$ denote the binomial coefficients for $l \leq m$. We obtain ,by ignoring terms of orders higher than three, that

$$\begin{aligned} \tau\partial_t\tilde{U}^o &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (e^{\tau\partial_t} - 1)^m \tilde{U}^o \\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} (P - H\partial_x + Q\partial_x^2 - R\partial_x^3 + \dots)^m \tilde{U}^o \\ &= \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} P^m \tilde{U}^o + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 P^{m-1} (-H\partial_x) \tilde{U}^o \end{aligned}$$

$$\begin{aligned}
& + \left\{ \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m} C_m^2 P^{m-2} H^2 \partial_x^2 + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 P^{m-1} Q \partial_x^2 \right\} \tilde{U}^o \\
& + \left\{ \sum_{m=3}^{\infty} \frac{(-1)^{m+1}}{m} C_m^3 P^{m-3} (-H \partial_x)^3 + \sum_{m=2}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 C_{m-1}^1 P^{m-2} (-H \partial_x) Q \partial_x^2 \right. \\
& \left. + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} C_m^1 P^{m-1} (-R \partial_x^3) \right\} \tilde{U}^o + \dots \\
& = \ln |P + 1| \tilde{U}^o - \frac{H}{1+P} \partial_x \tilde{U}^o + \left\{ -\frac{H^2}{2(1+P)^2} + \frac{Q}{(1+P)} \right\} \partial_x^2 \tilde{U}^o \\
& + \left\{ \frac{-H^3}{3(1+P)^3} + \frac{HQ}{(1+P)^2} - \frac{R}{1+P} \right\} \partial_x^3 \tilde{U}^o + \dots \\
& = \frac{1}{\tau} \ln \left| \frac{1-2q}{1+4\mu} \right| \tilde{U}^o - \frac{a}{2q-1} \partial_x \tilde{U}^o + \frac{h^2}{2\tau} \frac{(2q-1)(q+2\mu) - \nu^2 a^2 (1+4\mu)}{(2q-1)^2 (1+4\mu)} \partial_x^2 \tilde{U}^o \\
& + \frac{ah^2}{6} \frac{[(q+1)(2q-1) - 2\nu^2 a^2]}{(2q-1)^3} \partial_x^3 \tilde{U}^o + \dots \tag{26}
\end{aligned}$$

Thus we derive the modified equation for the oscillatory part:

$$\begin{aligned}
\partial_t \tilde{U}^o + \frac{a}{2q-1} \partial_x \tilde{U}^o &= \frac{1}{\tau} \ln \left| \frac{1-2q}{1+4\mu} \right| \tilde{U}^o \\
&+ \frac{h^2}{2\tau} \frac{(2q-1)(q+2\mu) - \nu^2 a^2 (1+4\mu)}{(2q-1)^2 (1+4\mu)} \partial_x^2 \tilde{U}^o \\
&+ \frac{ah^2}{6} \frac{(q+1)(2q-1) - 2\nu^2 a^2}{(2q-1)^3} \partial_x^3 \tilde{U}^o + \dots \tag{27}
\end{aligned}$$

Then, we write (27) as

$$\begin{aligned}
\partial_t \tilde{U}^o + \partial_x \tilde{U}^o - \varepsilon \partial_x^2 \tilde{U}^o &= \frac{1}{\tau} \ln \left| \frac{1-2q}{1+4\mu} \right| \tilde{U}^o + \frac{2a(q-1)}{2q-1} \partial_x \tilde{U}^o \\
&+ \frac{h^2}{2\tau} \frac{(2q-1)(q+2\mu) - [\nu^2 a^2 + 2\mu(2q-1)^2] (1+4\mu)}{(2q-1)^2 (1+4\mu)} \partial_x^2 \tilde{U}^o \\
&+ \frac{ah^2}{6} \frac{(q+1)(2q-1) - 2\nu^2 a^2}{(2q-1)^3} \partial_x^3 \tilde{U}^o + \dots \tag{28}
\end{aligned}$$

We call the zero term $\frac{1}{\tau} \ln \left| \frac{1-2q}{1+4\mu} \right| \tilde{U}^o$ in (28) a numerical damping term, which exerts dominant dissipation to suppress the oscillations caused by the relative phase error of high frequency modes. The first order term $\frac{2a(q-1)}{2q-1} \partial_x \tilde{U}^o$ can be used to estimate the extent of dispersion.

Now the numerical dissipation is not enough to control oscillations caused by the relative phase errors of high frequency modes, and only the numerical damping can suppress the oscillations. Thus, we need conditions $0 < \left| \frac{1-2q}{1+4\mu} \right| < 1$, which ensure the coefficient of the numerical damping term negative. Considering the stability condition of the scheme (9), we obtain $0 < \nu^2 a^2 \leq q + 2\mu < 1 + 4\mu$ and $q \neq \frac{1}{2}$.

4 Remarks and Conclusions

We use the method of modified equation analysis to obtain the mechanisms of dissipation and phase error of scheme (1.9)and The same to the modified equa-

tion analysis for scheme(8). **(1)** The numerical damping becomes stronger as q is closer to $\frac{1}{2}$. Particularly, if $q = \frac{1}{2}$, the oscillation is damped out immediately, by noting that $\lim_{q' \rightarrow \frac{1}{2}} \frac{1}{\tau} \ln \left| \frac{1-2q}{1+4\mu} \right| = -\infty$. So the numerical damping becomes infinite for $q = \frac{1}{2}$.

(2) In particular, if $q = 1$, the first order term vanishes; the numerical damping term is $-\frac{\ln|1+4\mu|}{\tau} \tilde{U}^o$, the numerical damping becomes stronger as μ is larger, and the numerical damping term will not vanish at all. when $1 - \nu^2 a^2 (1 + 4\mu) - 8\mu^2 > 0$, the numerical viscosity term still plays dissipation roles, but it is still weak in comparison with the numerical damping term.

References

1. Li, J., Yang, Z., Zheng, Y.: Characteristic decompositions and interactions of rarefaction waves of 2-D Euler Equations. *Journal of Differential Equations* 250, 782–798 (2011)
2. Li, J., Li, Q., Xu, K.: Comparison of the Generalized Riemann Solver and the Gas-Kinetic scheme for Compressible Inviscid Flow Simulations. *Journal of Computational Physics* 230, 5080–5099 (2011)
3. Dou, L., Dou, J.: The Grid: Time-domain analysis of lossy multiconductor transmission lines based on the LaxCWendroff technique. *Analog Integrated Circuits and Signal Processing* 68, 85–92 (2011)
4. Zhu, P., Zhou, S.: Relaxation Lax-Friedrichs sweeping scheme for static Hamilton-Jacobi equations. *Numer. Algor.* 54, 325–342 (2010)
5. Li, J., Liu, T., Sun, Z.: Implementation of the GRP scheme for computing spherical compressible fluid flows. *Journal of Computational Physics* 228, 5867–5887 (2009)
6. Li, J.-Q., Tang, H.-Z., Warnecke, G., Zhang, L.M.: Local Oscillations in Finite Difference Solutions of Hyperbolic Conservation Laws. *Mathematics of Computation*, S0025-5718(09)02219-4
7. Morton, K.W., Mayers, D.F.: Numerical Solution of Partial Differential Equations, 2nd edn. Cambridge University Press (2005)
8. Breuss, M.: The correct use of the Lax-Friedrichs method. *M2AN Math. Model. Numer. Anal.* 38, 519–540 (2004)
9. Breuss, M.: An analysis of the influence of data extrema on some first and second order central approximations of hyperbolic conservation laws. *M2AN Math. Model. Numer. Anal.* 39, 965–993 (2005)
10. Thomas, J.W.: Numerical Partial Differential Equations: Finite Difference Methods. Springer (1995)
11. Warming, R.F., Hyett, B.J.: The modified equation approach to the stability and accuracy of finite difference methods. *J. Comput. Phys.* 14, 159–179 (1974)
12. Tadmor, E.: Numerical viscosity and the entropy condition for conservative difference schemes. *Math. Comp.* 43, 369–381 (1984)

Fuzzy Time Series Forecasting Based on Weber-Fischna Law

Lihong Li, Yan Gao, and Yan Wang

School of Science, Hebei United University, Xin Hua Street 46, Tangshan,
063009, Hebei, P.R. China

Abstract. Time series model need lots of historical data, but the historical data are always incomplete, inaccurate and ambiguous. This uncertainty will affect the accuracy of forecasting. So the application of traditional forecasting model is limited. In this article, we apply the famous psychology law (Weber-Fischna law) in fuzzy time series models and predispose the historical data, so that we can fuzz data better. On the other hand, we use K-means cluster analysis to determine the center of class, which make the study more scientific and rigorous. In aspect of determining order, we find the best order by using the third criterion for judgment of order, which makes error less, makes precision of forecasting higher.

Keywords: Weber-Fischna law, Fuzzy Time Series, K-means cluster.

1 Introduction

Time series Forecasting is a method that historical data of forecasting target will be rank according to the time order, then its change trend will be analyzed, and we establish mathematical model to get quantitative Forecasting by extrapolation method. This kind of method is according to principle of coherence. At first, we suppose the trend of development for things yesterday and now is the same as it will be tomorrow. We get evolution law of the developing things in statistical data, and forecast the future developing trend for target. Time series model need lots of historical data, but the historical data are always incomplete, inaccurate and ambiguous. This uncertainty will affect the accuracy of forecasting. So the application of traditional forecasting model is limited. For this problem, Theory of fuzzy mathematics is used in course of forecasting, so that result of forecasting will be more conform to reality.

2 Fuzzy Time Series Analysis

2.1 Concept of Fuzzy Time Series

Song and Chissom (1993) proposed concept of Fuzzy Time Series for the first time: Suppose fuzzy set $f_i(t)$ ($i=1,2,\dots$) is defined in universe $Y(t)$ ($t=\dots,0,1,2,\dots$), and

$F(t)$ is the set of $f_i(t)$ ($i=1,2,\dots$), then $F(t)$ is called Fuzzy Time Series in $Y(t)(t=\dots,0,1,2,\dots)$. In this definition, $F(t)$ can be thought as 1inguistic variable, and $f_i(t)$ ($i=1,2,\dots$) are possible values for $F(t)$.

3 Current Study Situation for Fuzzy Time Series and Analysis

3.1 Current Study Situation for Fuzzy Time Series

Song and Chissom (1993) proposed concept of Fuzzy Time Series for the first time, which can dispose incomplete and fuzzy data in uncertain environment by using fuzzy logic theory. But in model of Song and Chissom, it needs lots of computing to determine fuzzy logic relationship, and it didn't consider the effect of interval gap in fuzzy set to forecasting precision. While in Sun's model, this aspect is considered. When internal gap is too big, the fuzzy time series lack volatility; when it is too small, the meaning of application of fuzzy time series can't be reflected.

3.2 Theory Basis of Improved Model

For the analysis, we propose a new forecasting method of fuzzy time series Based on psychology Weber-Fischna Law.

3.2.1 Weber-Fischna Law

Weber Law: $WB = \Delta I/I$, (in which WB is Weber score, I is original amount of stimulation, ΔI is increment of stimulation).

The formula tell us, when observing difference of two objects, we can detect the relative difference, not absolute difference, which is used to be verified in several sensory organ. Fischna Law: $S = M * \lg I$, (In which, S is feeling variable, M is constant, I is physical quantity.) Feeling intension will change as the logarithm of stimulation intension changes.

3.2.2 Determination of Threshold by Cluster Analysis

Refer to some literatures, we use K-means cluster method to determine class center, and use cluster center as threshold of different class. K-means arithmetic get the minimum sum of squared error for target function by using grad drop method. It has make convergence speed faster, so that we can solve the defects existing in average interval length for Song's method.

3.3 Improved Model for Fuzzy Time Series

This article combined Weber-Fischna law, proposes Fuzzy Time Series based on psychology law on the basis of Song and Chissom's model. Algorithms steps:

- 1). Collect data;
- 2). Repose of data, Repose the historical data based on Weber-Fischna law;
- 3). Define universe U; For all data, find out maximum value D_i and minimum D_{\min} ; Set two suitable value D_1 and D_2 , this two values are useful for team segment, then the range $U = [D_{\min} - D_2, D_{\max} - D_1]$ is obtained. In this article Universe $U = [D_{\min} - \sigma, D_{\max} + \sigma]$, σ is sample standard deviation.
- 4). Determine cluster number k, namely set length of discrete interval and membership degree;
- 5). K-means cluster analysis; Get cluster centre, Take adjacent cluster center $c_i, i = 1, 2, \dots, k$ as boundary points of subinterval in universe, namely threshold, then get each subinterval in universe, $A_1 = [D_{\min}, c_1]; A_2 = [c_1, c_2]; A_3 = [c_2, c_3]; A_4 = [c_3, c_4]; A_5 = [c_4, D_{\max}]$
- 6). Define fuzzy semantic value; semantic value mainly prefer to change degree of change value, so that we can define the fuzzy set which is belonged to according to the change degree.
- 7). Fuzzy up example data; put example data into suitable fuzzy set (according to change degree), namely, if real change rate is in the range of A_i at t period, it belong to fuzzy set of A. After data fuzzed, we can get each fuzzy set for which fuzzy time series F(t) belong to.
- 8). Calculate fuzzy logic relationship;
- 9). Defuzzification; in this article, we use center of gravity method to defuzzificate
- 10). Determine the best order; judgment of order, we use the third criterion for judgment of order.
- 11). Forecasting single value
- 12). Comparison of precisions for different forecasting models; apply each step into study of number of students registered in the beginning the new term. Then we compare the difference between studies of scholars before and this article. And we notice if the correct rate is higher than before.

4 Positive Study and Comparison of Precisions for Different Forecasting Models

Numbers of students actually enrolled in Alabama University from 1971 to 1992 is shown in table 1:

Table 1. Numbers of Students Actually Enrolled in Alabama University from 1971 to 1992

year	Number of students	year	Number of students
1971	13055	1975	15460
1972	13563	1976	15311
1973	13867	1977	15603
1974	14696	1978	15861

Based on Weber-Fischna law, we predispose the historical data, get the example data x_t , and $x_t = \frac{\lg(S_t) - \lg(S_{t-1})}{\lg(S_{t-1})}, t = 1972, 1973, \dots, 1992$.

4.1 Determining Universe U by Cluster Analysis

Define universe U. Universe $U = [D_{\min} - \sigma, D_{\max} + \sigma]$, in which, $D_{\min} = -0.00623$, $D_{\max} = 0.007524$, σ is sample standard deviation 0.003619589.

$$\therefore \text{Universe } U = [D_{\min} - \sigma, D_{\max} + \sigma] = [-0.00985, 0.011144]$$

Next, we take K-means clustering analysis, determine cluster centre, namely threshold.

We can classify the data into 4 classes by K—means algorithm, the last four cluster centers are:

$$c_1 = -0.0062255958355; c_2 = -0.002282175841876 \\ c_3 = 0.00114406485155; c_4 = 0.005525662868537$$

From the result of variance analysis, P-value is less than 0.05, it refers that the cluster result of K—means cluster algorithm is better.

Take adjacent cluster centre as threshold of subinterval in universe, then get every subinterval in universe,

$$A_2 = [-0.0062255958355, -0.002282175841875] \\ A_3 = [-0.002282175841875, 0.001144064851559] \\ A_4 = [0.001144064851559, 0.005525662868537] \\ A_5 = [0.005525662868537, 0.011144]$$

We can see, length of each subinterval is different.

4.2 Determination of Membership Degree

We first define semantic variable according to the analysis above:

A_1 Substantially descend, A_2 slightly descend, A_3 unchanging basically, A_4 slightly increase, A_5 substantially increase. Category which number of students enrolled every year belonged to is shown below partly:

Table 2. Category which Number of Students Enrolled Every Year Belonged to

Year	Ratio	Category	Year	Ratio	Category
1972	0.00401197	A_4	1976	-0.001005	A_3
1973	0.00232419	A_4	1977	0.001956629	A_4
1974	0.006051217	A_5	1978	0.00169569	A_4
1975	0.005254057	A_4			

We set frequency number in each example interval is $p_i (i=1,2,3,4)$. Then we have,

$$p_1 = 1; p_2 = 3; p_3 = 5; p_4 = 9; p_5 = 3.$$

By calculating, gravity of interval A_1 is $d_1 = -0.0062256$; gravity of interval A_2 is:

$$d_2 = \frac{-0.00328595 + (-0.00238701) + (-0.00245074)}{3} = -0.0027079$$

$$d_3 = 5.533E - 05; d_4 = 0.0036162734444; d_5 = 0.0065099036667$$

4.3 Calculating Fuzzy Logic Relationship and Defuzzification

As we use fuzzy composed calculating, orders must be above 2. Take 1989 for example, when second order is six, criterion matrix is composed by data of 1988, while calculating matrix is composed by data from 1981-1987. Then we get

$$F(1989) = (0, 0, 0, 0.5, 0.5).$$

By calculating, we can get

$$y(1989) = \frac{0.0036162734444 \times 0.5 + 0.0065099036667 \times 0.5}{0.5 + 0.5}$$

$$= 0.0050630885556$$

4.4 Determination of Best Order and Single Value Forecasting

We use the third criterion for judgment of order w : $E(w+1, w) = \frac{MSE(w+1)}{MSE(w)}$

We still take 1989 for example, when $w = 10$, forecasting value is 0.00554536, and

$$E(10, 9) = \frac{0.0055453602593}{0.0050630885556} = 1.0952524726,$$

From the rule above, the best order is nine. The best signal forecasting is

$$y(1989) = \frac{0.0036162734444 \times 0.5 + 0.0065099036667 \times 0.5}{0.5 + 0.5}$$

$$= 0.0050630885556$$

By calculating, we can see the comparison of forecasting and real values for number of students enrolled from 1974 to 1992 in figure 1.

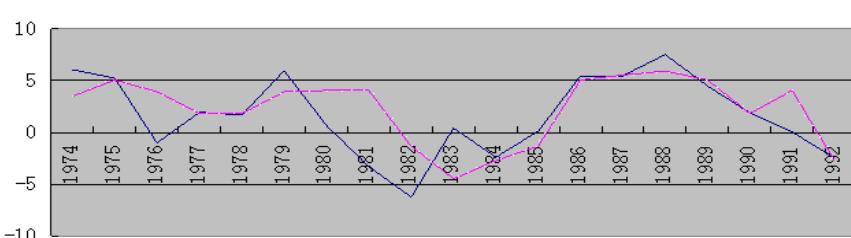


Fig. 1. Comparison of Forecasting and Real Values for Number of Students Enrolled from 1974 to 1992 (Solid line refer to real value*1000, the other line is fitted line)

4.5 Comparison of Precisions for Different Forecasting Models

At last, we compare the average equal square error for different fuzzy time series models, results are given in table 3.

Table 3. Average Equal Square Error for Different Fuzzy Time Series Models

Fuzzy time series method	Average error/%	Fuzzy time series method	Average error/%
Song-Chissom I	3.2	Markov method	2.6
Song-Chissom II	4.37	Hwang Method	3.18
Chen method	3.22	Based on Weber-Fischna law	2.745

From comparison analysis in table 3, we can see the precisions of fuzzy time series Based on Weber-Fischna law is higher than Song-Chissom, Chen, Hwang's models.

5 Conclusion and Expectation

In this article, we apply the famous psychology law (Weber-Fischna law) in fuzzy time series models and predispose the historical data, so that we can fuzz data better. On the other hand, we use K-means cluster analysis to determine the center of class, which make the study more scientific and rigorous. In aspect of determining order, we find the best order by using the third criterion for judgment of order, which makes error less, makes precision of forecasting higher. Study of this article has a lot of shortcomings. First of all, the number of examples is a little smaller, which will effect on the universality and representativeness. Second, we can only make superficial discussion because of our inadequate ability and pressing time.

Acknowledgments. Supported by: Nature Science Fund of Hebei province, (No. A 2012209030); Science research and development guide plan of Tang Shan (No. 111102033b).

References

- Pan, H.: Time Series Analysis. The Press of University of International Business Economics (January 2006)
- Hu, B.: Fuzzy theory basis. Wuhan University Press, Wuhan (2004)
- Yu, X., Ren, X.: Multivariate Statistical Analysis. China Statistics Press, Beijing (1999)
- Yang, Z.: Experimental psychology. Zhe Jiang Education Publishing House, Zhejiang (1998)
- Lin, H.: Forecasting research of stock market based on fuzzy time series. Master Dissertation of Beijing University of Technology, pp. 1–56
- Dan, H., Zhang, Y., Zhang, S.: A Kind of Modified K-mean Clustering Algorithm. Journal of Chongqing Technology and Business University: Natural Science Edition 26(2), 144–147 (2009)

7. Shen, B., Yao, M., Yi, W.-S.: Fuzzy time series analysis method based on least squares support vector machines. *Journal of Zhejiang University (Engineering Science)* 39(8), 1142–1147
8. Lin, Y., Yang, Y.: New Discussion of Forecasting Theory in stock market based on fuzzy time series. *Statistics and Decision* (8), 34–37 (2010)
9. Song, Q., Chissom, B.S.: Fuzzy time series and its models College of Education. *Fuzzy Sets and Systems* 54, 269–277 (1993)
10. Jilani, T.A., Muhammad, S., Burney, A.: A refined fuzzy time series model for stock market forecasting. *Science Direct*, 2857–2862
11. Shiva Raj Singh Department of mathematics, Bnaras Hindu University, Varaasi, A Simple time variant method for fuzzy time series forecasting India Taylor & Francis, pp. 305–321
12. Hao-Tien, Liu, N.-C., Wei, C.-G.: Improved time-variant fuzzy time series forecast *Fuzzy Optim Decis Making*, p. 4. Springer Science+Business Media, LLC (2009)

Oscillation of Runge-Kutta Methods for a Scalar Differential Equation with One Delay

Qi Wang¹, Jiechang Wen¹, and Fenglian Fu²

¹ School of Applied Mathematics, Guangdong University of Technology,
Guangzhou, 510006, China

² School of Environmental Science and Engineering, Guangdong University of Technology,
Guangzhou, 510006, China
bmwzwq@126.com

Abstract. Numerical oscillation of Runge-Kutta methods for differential equations with piecewise constant arguments is considered in this paper. The conditions of oscillation for the Runge-Kutta methods are obtained. It is proven that the numerical oscillation on the integer nodes are equivalent to the numerical oscillation on the any nodes and oscillation of the analytic solution is preserved by the Runge-Kutta methods. Moreover, the relationship between stability and oscillation is discussed for analytic solution and numerical solution, respectively. At last, several numerical simulations are carried out to support the theoretical analysis of the research.

Keywords: Runge-Kutta methods, Numerical solution, Oscillation, Stability.

1 Introduction

We are interested in the numerical oscillation of Runge-Kutta methods for the following differential equations with piecewise constant arguments (EPCA):

$$x'(t) = ax(t) + bx([t]), \quad x(0) = x_0, \quad (1)$$

where $a \neq 0, b, x_0$ are real constants and $[\cdot]$ denotes the greatest integer function.

In recent years, lots of authors have investigated the properties of the analytic solutions of EPCA (see [1-4] and the references therein). In particularly, oscillation of solutions of EPCA has received much attention (see [5-7] and the extensive bibliography therein). For further details on this subject, we refer the interested reader to the book by Wiener [8]. Nowadays, special interest has been devoted to the properties of numerical solution for EPCA, such as convergence [9], stability [10-12] and oscillation [13-15]. In this paper, we will consider numerical oscillation of Runge-Kutta methods for (1) which is different from the equations in [13-15] and discuss the relationship between stability and oscillation in theoretical and experimental sense.

We recall the definition of solution to (1) as follows. We say that a function $x(t)$ is a solution of (1) if the following conditions are satisfied:

- (i) $x(t)$ is continuous on $[0, \infty)$.
- (ii) The derivative $x'(t)$ exists at each point $t \in [0, \infty)$, with the possible exception of the points $[t] \in [0, \infty)$, where one-sided derivatives exist.
- (iii) (1) is satisfied on each interval $[n, n+1)$ for $n \in N$.

The following theorem gives the analytic solution of (1).

Theorem 1 (see [8]). (1) has on $[0, \infty)$ a unique solution

$$x(t) = m_0(\{t\}) b_0^{[t]} x_0, \quad (2)$$

where $\{t\}$ is the fractional part of t and

$$m_0(t) = e^{at} + (e^{at} - 1)a^{-1}b, \quad b_0 = m_0(1).$$

Theorem 2 (see [8]). The solution of (1) is asymptotically stable for all initial value x_0 if and only if

$$-\frac{a(e^a + 1)}{e^a - 1} < b < -a. \quad (3)$$

Definition 1. A non-trivial solution of (1) is said to be oscillatory if there exists a sequence $\{t_k\}_{k=1}^\infty$ such that $t_k \rightarrow \infty$ as $k \rightarrow \infty$ and $x(t_k)x(t_{k-1}) \leq 0$; otherwise it is called non-oscillatory. We say (1) is oscillatory if all the non-trivial solutions of (1) are oscillatory; we say (1) is non-oscillatory if all the non-trivial solutions of (1) are non-oscillatory.

Theorem 3 (see [8]). Every solution of (1) is non-oscillatory if and only if

$$b > \frac{ae^a}{1-e^a}.$$

By Definition 1, we can see that if a solution $x(t)$ of (1) is continuous and non-oscillatory, then it must be eventually positive or negative. That is, there exists a $\gamma \in R$ such that $x(t) > 0$ ($x(t) < 0$) for $t \geq \gamma$.

2 Numerical Oscillation and Non-oscillation

2.1 The Runge-Kutta Methods

Similar to [14], applying the v -stage Runge-Kutta methods (A, B, C) with $A = (a_{ij})v \times v$, $B = (B_1, B_2, \dots, B_v)^T$ and $C = (C_1, C_2, \dots, C_v)^T$ to (1) yield the recurrence relation

$$x_{km+l+1} = G(x)x_{km+l} + \frac{b}{a}(G(x) - 1)x_{km} \quad (4)$$

where $x = ha = a/m$, $l = 0, 1, \dots, m-1$ and $G(x) = 1 + xB^T(I - xA)^{-1}e$ is the stability function of the method. From [14], we know that the Runge-kutta methods preserve the original order for (1).

It is easily seen that (4) is equivalent to

$$x_{(k+1)m} = \left(G(x)^m + \frac{b}{a} (G(x)^m - 1) \right) x_{km} \quad (5)$$

and

$$x_{km+l+1} = \left(G(x)^{l+1} + \frac{b}{a} (G(x)^{l+1} - 1) \right) x_{km}, \quad 0 \leq l \leq m-2 \quad (6)$$

2.2 Oscillation Analysis

As to any given Runge-Kutta methods, we assume that $\delta_1 < 0 < \delta_2$ such that

$$0 < G(x) < 1, \text{ for } \delta_1 < x < 0, \quad 1 < G(x) < \infty, \text{ for } 0 < x < \delta_2,$$

which implies

$$0 < \frac{G(x)-1}{x} < \infty, \text{ for } \delta_1 < x < \delta_2.$$

Definition 2. A non-trivial solution $\{x_n\}$ of (6) is said to be oscillatory if there exists a sequence $\{n_k\}$ such that $n_k \rightarrow \infty$ as $k \rightarrow \infty$ and $x_{n_k} x_{n_k-1} \leq 0$; otherwise it is called non-oscillatory. We say (6) is oscillatory if all the non-trivial solutions of (6) are oscillatory; we say (6) is non-oscillatory if all the non-trivial solutions of (6) are non-oscillatory.

According to Definition 2, we can see that if a solution $\{x_n\}$ of (6) is non-oscillatory, then $\{x_n\}$ is eventually positive or negative.

Definition 3 (see [13]). We say the Runge-Kutta methods preserve the oscillation of (1) if (1) oscillates then there is a $h_0 > 0$ such that (6) oscillates for $h < h_0$. Similarly, we say the Runge-Kutta methods preserve the non-oscillation of (1) if (1) non-oscillates then there is a $h_0 > 0$ such that (6) non-oscillates for $h < h_0$.

In the following, we show the relationships of the non-oscillation between the integer nodes $\{x_n\}$ and the any nodes $\{x_{kn}\}$.

Theorem 4. $\{x_{kn}\}$ and $\{x_n\}$ are given by (5) and (6), respectively, then $\{x_n\}$ is non-oscillatory if and only if $\{x_{kn}\}$ is non-oscillatory.

Proof. If $\{x_n\}$ is non-oscillatory, then $\{x_{kn}\}$ must be non-oscillatory. Conversely, if $\{x_{kn}\}$ is non-oscillatory, without loss of generality, we assume that $\{x_{kn}\}$ is an eventually negative solution of (5), that is, there exists a $k_0 \in \mathbb{R}$ such that $x_{kn} < 0$

for $k > k_0$. We will prove $x_{km+l} < 0$ for all $k > k_0 + 1$ and $l = 0, 1, \dots, m-1$. Suppose $b < 0$, according to (6), if $a < 0$, then $0 < G(x) < 1$ and $G(x)^m \leq G(x)^l$, hence

$$x_{km+l} = \left(G(x)^l + \frac{b}{a} (G(x)^l - 1) \right) x_{km} \leq \left(G(x)^m + \frac{b}{a} (G(x)^m - 1) \right) x_{km} = x_{km+m} < 0.$$

If $a > 0$, then $1 < G(x) < \infty$ and $G(x)^{-m} \leq G(x)^{-l}$, so

$$\begin{aligned} G(x)^{-l} x_{km+l} &= \left(1 + \frac{b}{a} (1 - G(x)^{-l}) \right) x_{km} \\ &\leq \left(1 + \frac{b}{a} (1 - G(x)^{-m}) \right) x_{km} \\ &= G(x)^{-m} x_{km+m} < 0. \end{aligned}$$

So $x_{km+l} < 0$. This completes the proof.

By Theorem 4, we can get the following result.

Theorem 5. $\{x_{km}\}$ and $\{x_n\}$ are given by (5) and (6), respectively, then $\{x_n\}$ is oscillatory if and only if $\{x_{km}\}$ is oscillatory.

Theorem 6. (5) is oscillatory if and only if

$$b < \frac{aG(x)^m}{1 - G(x)^m}.$$

Proof. We can easily get this proof from the fact that (5) is oscillatory if and only if the corresponding characteristic equation has no positive roots.

Let

$$\Omega = \frac{ae^a}{1 - e^a}, \quad \Omega(m) = \frac{aG(x)^m}{1 - G(x)^m},$$

then we have the following lemma.

Lemma 1 $\Omega(m)$ and Ω satisfy three relations:

- (a) $\Omega(m) \rightarrow \Omega$ as $h \rightarrow 0$;
- (b) $\Omega \geq \Omega(m)$ if $e^x \geq G(x)$ for $a > 0$ or $e^x \leq G(x)$ for $a < 0$;
- (c) $\Omega < \Omega(m)$ if $e^x < G(x)$ for $a > 0$ or $e^x > G(x)$ for $a < 0$.

Proof. (b) If $a > 0$ and $e^x \geq G(x)$, then $e^a \geq G(x)^m$, which is equivalent to

$$\frac{1}{1 - e^a} \geq \frac{1}{1 - G(x)^m},$$

that is

$$\frac{e^a}{1 - e^a} \geq \frac{G(x)^m}{1 - G(x)^m},$$

so we have $\Omega \geq \Omega(m)$. We can prove the other cases in the same way.

From Theorems 3-6, we obtain the next result.

Theorem 7

- (a) The Runge-Kutta methods preserve the oscillation of (1) if and only if $\Omega < \Omega(m)$,
- (b) The Runge-Kutta methods preserve the non-oscillation of (1) if and only if $\Omega \geq \Omega(m)$.

Before the main theorems are obtained, we introduce a useful corollary.

Corollary 1 (see [10]). Suppose $G(z) = \varphi(z)/\psi(z)$ (where $\varphi(z), \psi(z)$ are polynomials) is the (r, s) -Padé approximation to e^z . Then

- (a) $G(x) < e^x$ if and only if s is even for all $x > 0$,
- (b) $G(x) > e^x$ if and only if s is odd for $0 < x < \xi$,
- (c) $G(x) > e^x$ if and only if r is even for all $x < 0$,
- (d) $G(x) < e^x$ if and only if r is odd for $\eta < x < 0$,

where ξ is a real zero of $\psi_s(z)$ and η is a real zero of $\varphi_r(z)$.

Applying Theorem 7, Lemma 1 and Corollary 1, we can get the following two theorems.

Theorem 8. Suppose $G(z)$ is the (r, s) -Padé approximation to e^z , the Runge-Kutta methods preserve oscillation of (1) if any of the following conditions is satisfied:

- (i) $a > 0, h < h_1$ and s is odd;
 - (ii) $a < 0, h < h_2$ and r is odd,
- where $h_1 = -\delta_1/a$, $h_2 = -\delta_2/a$.

Theorem 9. Suppose $G(z)$ is the (r, s) -Padé approximation to e^z , the Runge-Kutta methods preserve non-oscillation of (1) if any of the following conditions is satisfied:

- (i) $a > 0, h < h_1$ and s is even;
 - (ii) $a < 0, h < h_2$ and r is even,
- where $h_1 = -\delta_1/a$, $h_2 = -\delta_2/a$.

3 Relationship between Stability and Oscillation

Theorem 10. The numerical solution of (1) is asymptotically stable if and only if $-\frac{a(G(x)^m + 1)}{G(x)^m - 1} < b < -a$.

Proof. We can get this proof from the fact that $x_n \rightarrow 0$ as $n \rightarrow \infty$ if and only if $|\hat{\lambda}| < 1$, where

$$\hat{\lambda} = G(x)^m + \frac{b}{a} (G(x)^m - 1).$$

Setting

$$\Omega^* = -\frac{a(e^a + 1)}{e^a - 1}, \quad \Omega^*(m) = -\frac{a(G(x)^m + 1)}{G(x)^m - 1}.$$

By Theorems 2, 3, 6, 10 we can get the following theorems.

Theorem 11. The analytic solution of (1) is

- (i) oscillatory and unstable if $b \in (-\infty, \Omega^*)$;
- (ii) oscillatory and asymptotically stable if $b \in (\Omega^*, \Omega)$;
- (iii) non-oscillatory and asymptotically stable if $b \in (\Omega, -a)$;
- (iv) non-oscillatory and unstable if $b \in (-a, +\infty)$.

Theorem 12. The numerical solution of (1) is

- (i) oscillatory and unstable if $b \in (-\infty, \Omega^*(m))$;
- (ii) oscillatory and asymptotically stable if $b \in (\Omega^*(m), \Omega(m))$;
- (iii) non-oscillatory and asymptotically stable if $b \in (\Omega(m), -a)$;
- (iv) non-oscillatory and unstable if $b \in (-a, +\infty)$

4 Numerical Experiment

We consider the following four problems:

$$x'(t) = x(t) - 2x([t]), \quad x(0) = 1, \quad (7)$$

$$x'(t) = -x(t) - 3x([t]), \quad x(0) = 1, \quad (8)$$

$$x'(t) = 2x(t) - 2.1x([t]), \quad x(0) = 1, \quad (9)$$

$$x'(t) = -3x(t) + 4x([t]), \quad x(0) = 1. \quad (10)$$

The analytic solutions of (7) and (8) are oscillatory; the analytic solutions of (9) and (10) are non-oscillatory according to Theorem 3. In Figs. 1-2, we draw the figures of the analytic solutions and the numerical solutions by using 1-Gauss-Legendre method, 2-Radau IA method and 2-Lobatto IIIC method. From the two figures, we can see that the numerical solutions of (7) and (8) are oscillatory; the numerical solutions of (9) and (10) are non-oscillatory, which are in agreement with Theorems 8 and 9.

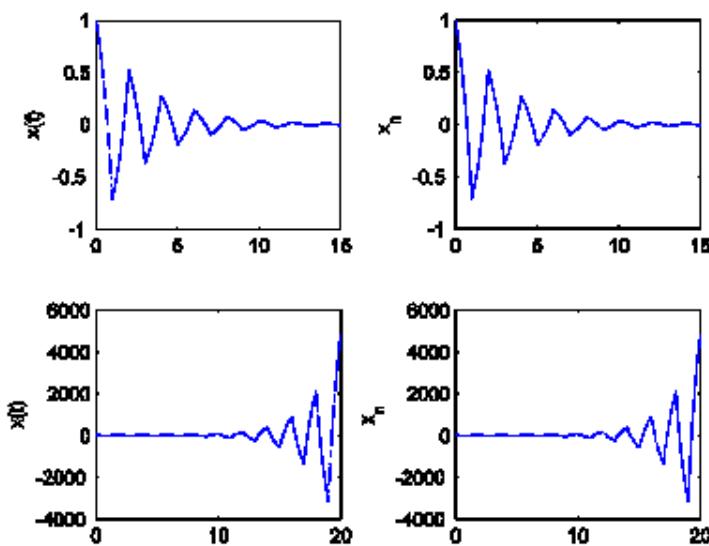


Fig. 1. The analytic solution and the numerical solution with $m=50$. 1-Gauss-Legendre method for (7) (First row) and 2-Radau IA method for (8) (Second row).

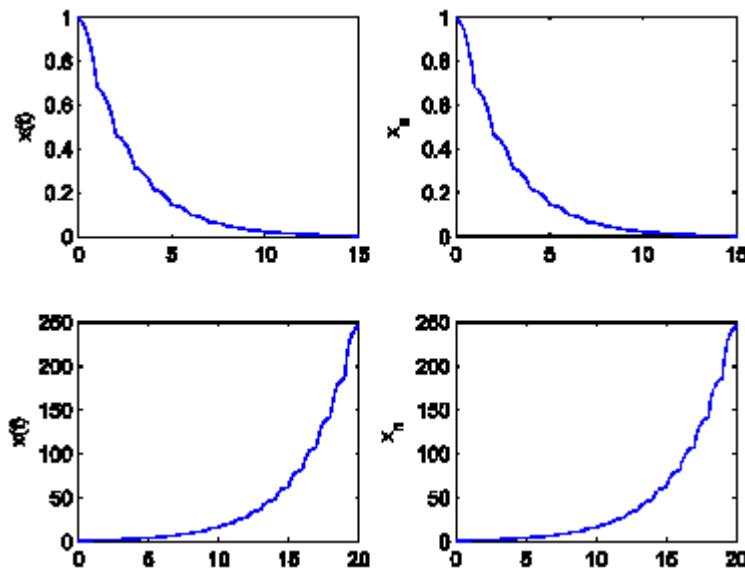


Fig. 2. The analytic solution and the numerical solution with $m=50$. 2-Radau IA method for (9) (First row) and 2-Lobatto IIIC method for (10) (Second row).

In the first row of Fig. 1, put $m = 50$, we have that $\Omega \approx -1.5820$, $\Omega^* \approx -2.1640$, $\Omega(m) \approx -1.5819$ and $\Omega^*(m) \approx -2.1639$. Obviously, $b = -2 \in (\Omega^*, \Omega)$ and $b = -2 \in (\Omega^*(m), \Omega(m))$. Therefore, the analytic solutions and the numerical solutions of (7) are both oscillatory and asymptotically stable according to Theorems 11 and 12, which are coincide with Fig. 1. For (8)-(10), we can test them in the same way (see Figs. 1-2).

5 Conclusions

In this article we have analyzed numerical oscillation of an important EPCA and the relationship between stability and oscillation is also presented. We conclude from the results that the Runge-Kutta methods are applicable for the study of EPCA. Currently, we are working on extending this research to the multidimensional and complex coefficient problems.

Acknowledgments. Many thanks to Professor Mingzhu Liu for his valuable suggestions which improve the paper significantly. This work is supported by National Nature Science Foundation of China (No. 51008084).

References

1. Akhmet, M.U., Arugaslan, D., Yilmaz, E.: Stability in Cellular Neural Networks with a Piecewise Constant Argument. *J. Comput. Appl. Math.* 233, 2365–2373 (2010)
2. Akhmet, M.U., Arugaslan, D., Yilmaz, E.: Method of Lyapunov Functions for Differential Equations with Piecewise Constant Delay. *J. Comput. Appl. Math.* 235, 4554–4560 (2011)
3. Dimbour, W.: Almost Automorphic Solutions for Differential Equations with Piecewise Constant Argument in a Banach Space. *Nonlinear Anal.* 74, 2351–2357 (2011)
4. Akhmet, M.U., Arugaslan, D., Yilmaz, E.: Stability Analysis of Recurrent Neural Networks with Piecewise Constant Argument of Generalized Type. *Neural Networks* 23, 805–811 (2010)
5. Fu, X.L., Li, X.D.: Oscillation of Higher Order Impulsive Differential Equations of Mixed Type with Constant Argument at Fixed Time. *Math. Comput. Model.* 48, 776–786 (2008)
6. Luo, Z.G., Shen, J.H.: New Results on Oscillation for Delay Differential Equations with Piecewise Constant Argument. *Comput. Math. Appl.* 45, 1841–1848 (2003)
7. Shen, J.H., Stavroulakis, I.P.: Oscillatory and Nonoscillatory Delay Equations with Piecewise Constant Argument. *J. Math. Anal. Appl.* 248, 385–401 (2000)
8. Wiener, J.: Generalized Solutions of Functional Differential Equations. World Scientific, Singapore (1993)
9. Song, M.H., Liu, X.: The Improved Linear Multistep Methods for Differential Equations with Piecewise Continuous Arguments. *Appl. Math. Comput.* 217, 4002–4009 (2010)
10. Liu, M.Z., Ma, S.F., Yang, Z.W.: Stability Analysis of Runge-Kutta Methods for Unbounded Retarded Differential Equations with Piecewise Continuous Arguments. *Appl. Math. Comput.* 191, 57–66 (2007)

11. Song, M.H., Yang, Z.W., Liu, M.Z.: Stability of θ -Methods for Advanced Differential Equations with Piecewise Continuous Arguments. *Comput. Math. Appl.* 49, 1295–1301 (2005)
12. Dai, H.Y., Liu, M.Z.: Mean-square Stability of Stochastic Differential Equations with Piecewise Continuous Arguments. *Heilongjiang Univ. J. Nat. Sci.* 25, 625–629 (2008)
13. Liu, M.Z., Gao, J.F., Yang, Z.W.: Oscillation Analysis of Numerical Solution in the θ -Methods for Equation $x'(t)+ax(t)+a_1x([t-1])=0$. *Appl. Math. Comput.* 186, 566–578 (2007)
14. Liu, M.Z., Gao, J.F., Yang, Z.W.: Preservation of Oscillations of the Runge-Kutta Method for Equation $x'(t)+ax(t)+a_1x([t-1])=0$. *Comput. Math. Appl.* 58, 1113–1125 (2009)
15. Gao, J.F., Liu, M.Z.: Numerical Oscillations of the θ -Method for Advanced Delay Differential Equations with Piecewise Continuous Arguments. *Heilongjiang Univ. J. Nat. Sci.* 25, 196–203 (2008)

Analysis of Wall Criteria in Numerical Computation of Shock Wave/Layer Interaction

Xinwen Xu and Yong Yu

School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China
yuyong@bit.edu.cn

Abstract. A new method, wall criteria, which evaluates numerical result satisfying interacting shear flow theory, is presented. This method can replace the analysis of grid convergence. Two ways to simplify two-dimensional Navier-Stokes equations are introduced: direct simplification and ignoring simplification from interacting shear flow theory. Then, an investigation of shock wave/layer interaction on a flat plate was studied through numerical simulation. The comparison of every term's magnitude in the expressions from direct method verifies the rationality and accuracy of the expressions deduced from the ignoring method. Lastly, the first and third wall criterion have been proved reasonable. Two nondimensionalized numbers B_c and E_c were introduced to estimate the quality of numerical outcome.

Keywords: shock wave/layer interaction on flat plate, wall criteria, criterion formula, numerical validation.

1 Introduction

Depending on the spectacular development of computer technology during the past decades, computational fluid dynamics (CFD) has constituted the “third approach” in the research of fluid dynamics from a local tentative alternative for experiments that cannot conduct initially. Nowadays, CFD has been used widely in science research and engine practice such as astronavigation, energy, meteorology environment, etc. However, the CFD development people are always confused by the verification and accuracy of numerical simulation [1]. Is the outcome generated from the code on computer correct or accurate enough? If the result is worth trusting, how can we assess its authentic extent? Researchers proposed various validation measures to evaluate the numerical consequence, mainly including experimental contrast, accurate solution contrast, artificial solution contrast, grid convergence analysis and so on. When analyzing a special flow with numerical method, most of the assessing means are not acquired easily except the grid convergence analysis [2]. Can we find a new way to verify the reliability and accuracy of the numerical result? GAO [3-6] explored a novel method named wall criteria which assess the numerical outcome on no-slip wall through the characteristic of interacting shear flow (ISF) theory and near surface complex ISF theory. This new criterion has attracted more attention and may be applied as a validation mechanism.

2 Two Different Simplifying Means for Navier-Stokes Equations

The two-dimension momentum and energy equations in conservation form are given in the following expressions:

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} = -\frac{\partial p}{\partial x} + \frac{\partial}{\partial x} \left(\lambda \nabla \cdot \mathbf{V} + 2\mu \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y} \left[\mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right] \quad (1)$$

$$\frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho v^2)}{\partial y} + \frac{\partial(\rho uv)}{\partial x} = -\frac{\partial p}{\partial y} + \frac{\partial}{\partial y} \left(\lambda \nabla \cdot \mathbf{V} + 2\mu \frac{\partial v}{\partial y} \right) + \frac{\partial}{\partial x} \left[\mu \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right) \right] \quad (2)$$

$$\rho \frac{De}{Dt} = \rho \dot{q} + \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) - p \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \lambda \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right)^2 + \mu \left[2 \left(\frac{\partial u}{\partial x} \right)^2 + 2 \left(\frac{\partial v}{\partial y} \right)^2 + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right)^2 \right] \quad (3)$$

The main idea of ISF theory proposed by Gao is: near wall ISF includes two parts of viscous and inviscid flows. In this local area, convection-dominant in the tangential direction of the wall surface and convection-diffusion competes in the normal.

Associating ISF theory and no-slip wall, three sets of ISF equations (6)-(8) can be obtained from the simplifying Navier-Stokes equations. For the sake of writing convenience, the equations are called wall criterion I, wall criterion II and wall criterion III (energy criterion) respectively. The three expressions are written as:

$$\text{wall criterion I: } -\frac{\partial p}{\partial x} + \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} \right) \cong 0 \quad (4)$$

$$\text{wall criterion II: } -\frac{\partial p}{\partial y} + \frac{\partial}{\partial x} \left(\mu \frac{\partial v}{\partial x} \right) \cong 0 \quad (5)$$

$$\text{wall criterion III: } \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + \mu \left(\frac{\partial u}{\partial y} \right)^2 \cong 0 \quad (6)$$

Enlightened from the deduced progress of wall criteria, author simplifies Navier-Stokes equations directly associating the characteristics of shock wave/layer interaction rather than ISF theory to deduce the following expressions (7)-(9) called criterion formulas temporarily. Besides, every term in the criterion formula is labeled below the expression for the later convenience of discussion.

$$\text{criterion formula I: } -\frac{\partial p}{\partial x} + \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial y} \right) - \frac{2}{3} \frac{\partial}{\partial x} \left(\mu \frac{\partial v}{\partial y} \right) + \frac{\partial}{\partial y} \left(\mu \frac{\partial v}{\partial x} \right) = 0 \quad (7)$$

BC_ISF1 BC_ISF2 BC_VAR1 BC_VAR2

$$\text{criterion formula II: } -\frac{\partial p}{\partial y} + \frac{4}{3} \frac{\partial}{\partial y} \left(\mu \frac{\partial v}{\partial y} \right) + \frac{\partial}{\partial x} \left(\mu \frac{\partial u}{\partial y} \right) - \frac{2}{3} \frac{\partial}{\partial y} \left(\mu \frac{\partial u}{\partial x} \right) = 0 \quad (8)$$

NC_ISF1 NC_ISF2 NC_VAR1 NC_VAR2

$$\text{criterion formula III: } \frac{\partial}{\partial y} \left(k \frac{\partial T}{\partial y} \right) + \mu \left(\frac{\partial u}{\partial y} \right)^2 + \frac{\partial}{\partial x} \left(k \frac{\partial T}{\partial x} \right) - p \frac{\partial v}{\partial y} + \frac{4}{3} \mu \left(\frac{\partial v}{\partial y} \right)^2 = 0 \quad (9)$$

EC_ISF1 EC_ISF2 EC_VAR1 EC_VAR2 EC_VAR3

Obviously, the criterion formulas that are from direct simplification are slightly complicated than wall criteria from ISF theory. It suggests former are more accurate than later as the former doesn't ignore any term.

As the simple geometrical construction and representative flow characteristics of shock wave/layer interaction on a flat plate, it was studied with numerical computation in the following content. Then, the rationality and accuracy of wall criteria was discussed and verified.

3 Numerical Simulation and Comparison with Experimental Data

Here, analysis and simulation of the shock wave/layer interaction focus on the effect of grid scales to the grid convergence rather than Mach number, incident angle, turbulence, Reynolds number, three-dimension effect or numerical scheme to pressure distribution and heat transfer[7-10]. The free stream parameters refer the experiment condition, which was conducted by G. DEGREZ[11].

Detailed parameters in numerical computation are $Ma=2.15$, $Re=0.96 \times 10^5/m$, static temperature $T=157K$, shock wave angle $\theta=30.8^\circ$, the length of plate in computational domain $L=0.16M$. The laminar model and second order accurate space discretization of Ros's FDS are applied in numerical computation. Boundary condition in the computational domain was given as the following figure 1.

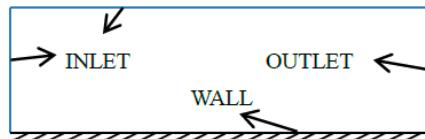


Fig. 1. Boundary conditions in computational domain

When incident oblique shock wave meets layer on the plate, there is a separation bubble near the interaction area. CFD program and author's post-processing procedure are sensitive to scales of the normal grid near the wall since the layer and bubble thickness depend on the grid scale largely.

Table 1. Various grid scales adopted in computation

case	A	B	C	D	E
X(mm)	3.75	2.0	1.0	0.5	0.25
Y(mm)	0.75	0.4	0.2	0.1	0.05

For the better accuracy of separation bubble and less computation cost, we divide the normal domain into average zone which is near the wall (height=6mm) and geometric growth zone which is above the average zone. Along flow direction, average grid also was employed. Detailed scales of various grid density in average zone were given in table 1.

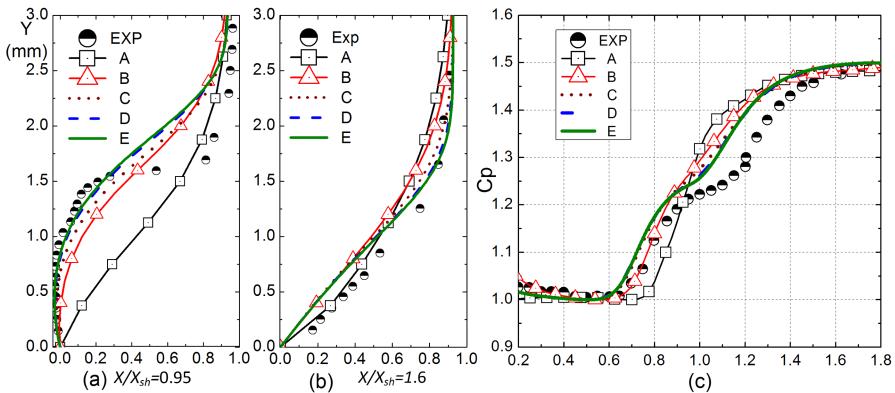


Fig. 2. Experimental and computed comparison of velocity profiles at two normal locations in boundary-layer and distribution of pressure coefficient along wall surface in various grid scales

Figure 2(a) shows the velocity profiles when y is below 1.5mm and over 2.7mm agreeing well with experiment as halving the grid at $X/X_{sh}=0.95$. But the smooth shifting in the middle segment maybe deduce by the slightly high artificial viscosity. Figure 2(b) reports the numerical solution approaching the experiment gradually when y is beyond 1.5mm as the grid refinement at the location of $X/X_{sh}=1.6$. However, the over prediction near the wall may induce by ignoring span-width in the three-dimension experiment or grid span larger than velocity gradient. The pressure along the wall has been normalized by the minimum pressure and its coefficient was plotted in figure 2(c). The grid convergence has been arrived from comparing figures with the grid refinement. But the reverse flow region and neighboring area, over prediction was seen maybe due to ignoring the effect of three-dimension, too.

4 Order of Magnitude Analysis for the Criterion Formulas

According to the previous comparison between numerical outcome in various grid density and experimental data, numerical computation can be satisfied when the grid scale equals case D. The following order of magnitude analysis uses second order accuracy difference to discretize every term in expression (7)-(9) basing on the scale of case D.

Figure 3(a) shows clearly BC_VAR terms that are extremely small compared with BC_ISF terms in criterion formula I. And they stay almost zero along the wall surface. These errors which are introduced through ignoring tiny terms in criterion formula I can be accepted for a correct estimating criterion. As a result, wall criterion I is appropriate to evaluate a numerical result coinciding the features of ISF flow.

For the criterion formula II, it's prevalent to set $\partial p/\partial y=0$ along normal direction in most CFD programs. Consequently, it would introduce unacceptable errors and even lead to an incorrect result without the NC_ISF1 in criterion formula II. Hence, it's unreliable to assess a numerical result with wall criterion II.

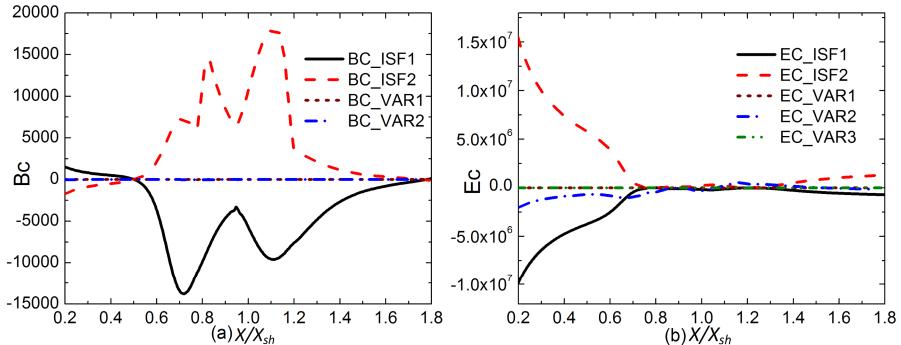


Fig. 3. Comparison of various terms' order of magnitude in criterion formula I and III

As shown in figure 3(b), the wall criterion III is neither best nor worst one compared with wall criterion I and II, as the wall criterion III discards EC_VAV2(pressure component) which is relatively small when X/X_{sh} is smaller than 0.6 or greater than 1.4. It's rational to abandon EC_VAR terms during this region. But EC_VAR2 competes with EC_ISF1 and EC_ISF2 in the similar order of manitude near the area when X/X_{sh} is between 0.6 and 1.4. It will bring major errors without EC_VAV2 in this area.

As a consequence, the wall criterion I discarding two terms which approach zero from criterion formula I is a perfect evaluating criterion. Similarly, wall criterion III which ignored EC_VAR terms is agreement with the criterion formula III except the middle section. Therefore, the wall criterion III is somewhat an acceptable criterion for evaluation. In contrast, the wall criterion II is recognized as the worst one.

5 Evaluation with Wall Criteria for Previous Numerical Result

Although there are some errors to the third wall criterion near the median zone, it and wall criterion I deduced from ISF theory still deserve to be tried as effective means to estimate the numerical result when lacking experimental data and grid comparisons. However, how to estimate the numerical outcome with these wall criteria? The detailed progress will be given in the following content.

The sum of BC_{ISF1} and BC_{ISF2} is a relatively small value from the former analyzing of order of magnitude. Its value which approaches zero further reflects a preferable and reliable numerical result. Likewise, energy criterion is in the same situation. Then, the sums of two terms of the *wall criterion I* and *III* were nondimensionalized by $\rho u^2/X_l$ and $\rho u^3/X_l$ respectively. This statement can be expressed as following expression (10):

$$Bc = \frac{-\frac{\partial p}{\partial x} + \frac{\partial}{\partial y}(\mu \frac{\partial u}{\partial y})}{\frac{\rho u^2}{X_l}} , \quad Ec = \frac{\frac{\partial}{\partial y}(k \frac{\partial T}{\partial y}) + \mu \left(\frac{\partial u}{\partial y} \right)^2}{\frac{\rho u^3}{X_l}} \quad (10)$$

Here, ρ and u are density and velocity of free stream from the experiment. X_l is a characteristic length, and $X_l=X_{sh}=0.08M$.

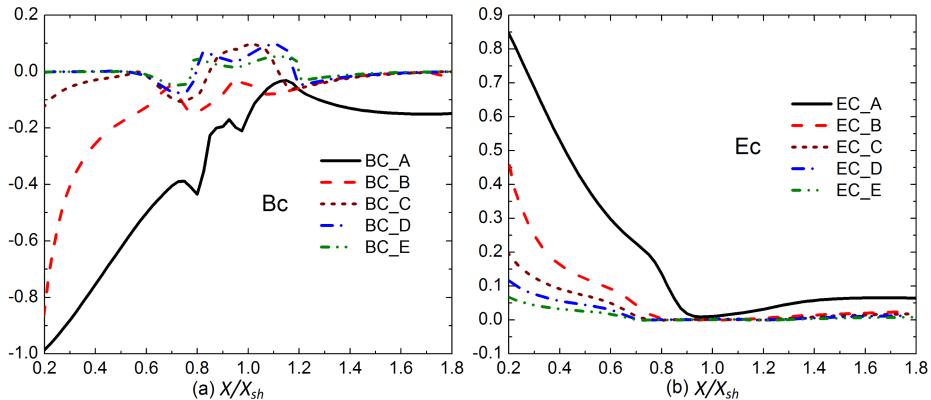


Fig. 4. Varying tendency of nondimensionalized number E_c and S_c with grid refinement

Author processed the previous five sets of numerical result of various scales with the post-processing procedure. Then, the varying tendency of nondimensionalized number with halving the grid was plotted in the following figures 4(a) and 4(b).

As shown distinctly in figures 4(a) and 4(b), nondimensionalized numbers B_c and E_c approach zero farther with the grid refinement. At the same time, numerical result agrees well with experimental data as B_c and E_c access gradually zero. For this special flow, grid convergence has been satisfied for the case 4. If only one set grid was supplied, there would be a relatively accurate numerical result when B_c is less than 0.1 in figure 4(a) or E_c is below 0.15 in figure 4(b).

6 Conclusion

An investigation of an oblique incident shock wave/layer interaction on a flat plate was studied with numerical simulation. Then, the numerical outcome was processed by simplified Navier-Stokes equations directly to verify the rationality of three wall criteria from ISF theory.

The wall criterion I in flow direction can be treated as a perfect evaluating criterion since the ignoring terms are extremely tiny. Similarly, the energy criterion is relatively reasonable for discarding terms are small in most area except the middle area on the wall. By contrast, the second wall criterion is the worst one because of the unreasonable ignoring terms.

Two nondimensionalized numbers B_c and E_c were introduced for estimating the numerical result. The distance from zero to the nondimensionalized value indicates the quality: A shorter distance means a better numerical result.

References

1. Yan, C., Xi, K., et al.: Review of the Drag Prediction Workshop Series 41, 776–784 (2011)
2. Yu, Y.: New Cfd Validation Method with Application to Verify Computation of Near Wall Flow. Journal of Beijing Institute of Technology 19, 259–263 (2010)

3. Gao, Z.: Viscous-Inviscid Interaction Flow Theory. *Acta Mechanica Sinica* 22, 8–19 (1990) (in Chinese)
4. Gao, Z.: The Wall-Surface Criteria with Application to Evaluating Creditability of CFD Simulations. *Acta Aerodynamica Sinica* 26, 378–393 (2008) (in Chinese)
5. Gao, Z.: Interacting Shear Flows (ISF) and Boundary Layer Flows and Application of ISF Theory to Computational Fluid Dynamics. *Advance in Mechanics* 38, 114–116 (2008) (in Chinese)
6. Gao, Z.: Invariance of Interactive-Strcture Between Convection and Diffusion. *Acta Mechanica Sinica* 21, 13–22 (2007) (in Chinese)
7. Wang, G., Jiang, X., Xue, R.: Study On Interaction Between Oblique Shock Wave and Flat Laminar Boundary Layer. *Journal of Projectiles, Rockets, Missiles and Guidance* 28, 206–209 (2008) (in Chinese)
8. Talbot-Stern, J.L., Auld, D.J.: Direct Simulation (MONTE CARLO) of Oblique Shock Wave-Laminar Boundary Interactions, AIAA-1998-2756
9. Benay, R., Chanetz, B., Délery, J.: Code Verification/Validation with Respect to Experimental Data Banks. *Aerospace Science and Technology* 7, 239–262 (2003)
10. Fedorova, N.N., Fedorchenko, I.A.: Computations of Interaction of An Incident Oblique Shock Wave with A Turbulent Boundary Layer on A Flat Plate. *Journal of Applied Mechanics and Technical Physics* 3, 358–366 (2004)
11. Degrez, G., Boccadoro, C.H., Wendt, J.F.: The Interaction of An Oblique Shock Wave With A Laminar Boundary Layer Revisited. An Experiment and Numerical Study. *J. Fluid Mech.* 177, 247–263 (1987)

A Voxelization Algorithm for 3D Body-Centered Cubic Line Based on Adjunct Parallelepiped Space

Lijun He, Yongkui Liu, Degao Wang, and Jian Yun

School of Computer Science and Engineering, Dalian Nationalities University,
Dalian 116600, China
helijun2002@hotmail.com

Abstract. A new voxelization algorithm is proposed for drawing lines on 3D body-centered cubic (BCC) grid, a kind of honeycomb space. Its main contribution is to apply the 3D Bresenham algorithm, a popular algorithm for drawing 3D lines on a cubic grid, to produce the BCC grid occupied by 3D lines, with the help of adjunct parallelepiped space, having the same center and basis vectors with the BCC grid. The adjunct parallelepiped line is easy to generate using the existed 3D cubic Bresenham algorithm. Because of the one-to-one correspondence between the parallelogram cells of parallelepiped space and the voxels of the BCC space, then the 3D BCC line voxelization is generated. This procedure is characterized by a simple discriminator. A derivation for this discriminator given in the paper confirms that all calculations can be realized using only integer arithmetic.

Keywords: Body-Centered Cubic (BCC) Grid, Voxelization, Algorithm, Voxel, Line.

1 Introduction

The development of volume graphics has lead to growing interest in the 3D line voxelization algorithms that convert a 3D continuous line into a discrete line representation. Being basic but very important in volume graphics, these algorithms are used for synthesizing voxel-based objects in volume graphics as the 3D line itself is also used to build block for voxelizing more complex objects. The 3D line voxelization algorithms are also employed for ray traversal in voxel space[1, 2]. It is extended from Bresenham's algorithm [3] to three-dimensional straight lines [4]. Similarly, an assortment of 3D scan-conversion algorithms [5] is introduced to scan-convert 3D geometric objects into their discrete voxel-map representation within the cubic grid. Then a tripod algorithm which generates an exact 3D discrete line on the cubic grid [6] is considered. Furthermore, an algorithm for naive 3D discrete line drawing [7] is given. An algorithm based on a linear incremental algorithm which recognizes any set of points on the cubic grid [8] is proposed. More recent studies [9] propose a volumetric drawing system that directly extracts and renders linear features that lay on isosurfaces within the volume.

Although the cubic grid seems to be more natural, alternatives to the cubic grid in 3D have also been considered. Two honeycomb graphical models in which the voxels

are hexagonal prisms are introduced in [10]. Three-dimensional grid where the voxels are rhombic dodecahedra and truncated octahedron are called FCC (face-centered cubic) grid and BCC (body-centered cubic) grid, respectively. Here voxel is defined as voronoi region of the grid and in this paper voxel is always refers to the voronoi region of the grid. The BCC grid and FCC grid are the three-dimensional equivalents of the two-dimensional hexagonal grid. Many recent studies have focused on BCC grid and FCC grid in [11-14]. Some properties and advantages of grid based on BCC grid and FCC grid have been studied, e.g., in [15]. Moreover, an algorithm for computing surface skeletons on the BCC grid and the FCC grid is presented in [16].

The BCC grid is the three-dimensional equivalents of the two-dimensional hexagonal grid in the sense of the favorable volumetric sampling pattern due to its optimal spectral sphere packing property. As the BCC grid has advantages above, some researchers investigated BCC grid in more detail and adapted several volume rendering methods to it. Carr et al [17] proposed marching octahedral, modified marching octahedral and modified marching hexahedra to generate isosurfaces on the BCC grid. Theußl et al.[18] implemented splatting on BCC grid. Sweeney et al adapted the shear-warp algorithm to the BCC grid. Dornhofer modified Fourier Domain Volume Rendering for use on the BCC grid. Ibáñez et al. A vectorial algorithm [19] is established to trace discrete straight lines on nonorthogonal grids in any dimension including the BCC grid. Ray casting algorithm provides results of very high quality, usually it is considered to provide the best image quality with the slowest speed. So it is very important to accelerate the ray casting algorithm to derive the high quality of image. The aim of this paper is to give a more efficient line voxelization algorithm on the BCC grid which can speed optimization of volumetric ray casting process. The algorithm is based on the adjunct parallelepiped grid and the 3D cubic Bresenham's line drawing algorithm. The contribution of the new algorithm is that up to three voxels can be produced in a step. As will be shown, the algorithm can be implemented using integer representation only. In this way, it is faster, and the accumulation of rounding errors is eliminated completely.

At first, we will briefly consider BCC grid in section 2. The adjunct parallelepiped space of the BCC system and the correspondence between them are introduced in section 3. The line voxelization algorithm is proposed on 3D BCC grid in detail in section 4. The experimental results are considered in section 5 and the paper is concluded in section 6.

2 BCC (Body-Centred Cubic) Grid

It is known that \mathbb{R}^3 can be uniformly tiled by parallelepipeds, hexagonal prisms, rhombic dodecahedrons, or truncated octahedrons. The corresponding three-dimensional grid where the voxels are truncated octahedrons is called the BCC (body-centered cubic) grid.

A BCC grid can be constructed from a pre-existing cubic grid. The following procedure is just one possible way of constructing a BCC grid, and is by no means the only one. Starting with a unit cubic grid, a second identical grid is added with a translation of (0.5, 0.5, 0.5), so that the new grid nodes fall into the center of the

original cubic voxels. The Voronoi regions so called voxels around all of these grid nodes have the shape of a truncated octahedron. Consider a discrete space \mathbf{N} consisting of slices composed by truncated octahedron, consecutively placed one by one. In contrast to the case of cubic grid, voxels here vertex-adjacency and edge-adjacency are both impossible. Thus, a connected discrete object is always face-adjacent and tunnel-free.

The voxels form a voxelized space \mathbf{N} based on the BCC grid. A coordinate system is defined on space \mathbf{N} , as follows. The coordinate system is simply one choice among many other possible coordinate systems. A voxel is chosen and its center O is defined to be the origin of the coordinate system. The origin's coordinates are all zeros, i.e., $O = (0, 0, 0)$. Next three coordinate axes OX , OY and OZ are fixed as shown in Fig 1, where OX , OY and OZ are “non-orthogonal” among themselves. And that OX , OY are orthogonal to adjacent two hexagons of the chosen truncated octahedron (voxel), while OZ is orthogonal to the square face that is adjacent to the above two hexagonal faces. The basis vectors u , v and w of the coordinate system are aligned with the coordinate axes, where the vector u is parallel to the OX axis and it connects the center of the origin voxel to the center of the adjacent voxel through the hexagonal face that is orthogonal to the OX axis. The vectors v and w are as similar as vector u . It must be point out that the vectors u , v , w are not unit vectors. In particular, $|u|= \sqrt{3}/2$, $|v|= \sqrt{3}/2$, $|w|=1$.

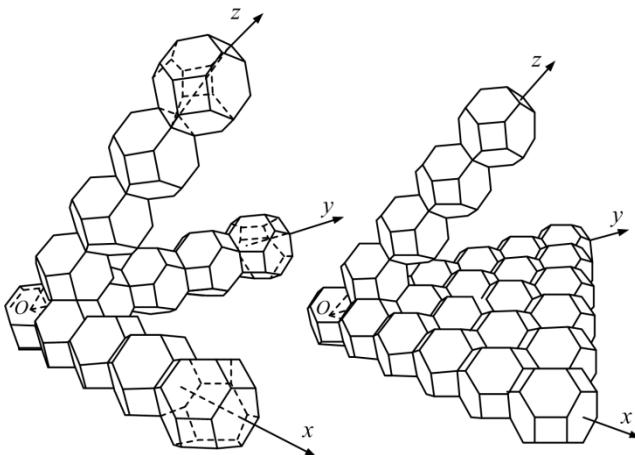


Fig. 1. The axes of the coordinate system $Oxyz$

The centers of the voxels of the BCC space form a lattice L , as u , v and w form a basis of L . The BCC coordinate system thus defined is denoted by $OXYZ$. The coordinate axes OX , OY and OZ define three planes by pairs: the plane OXY (defined by the axes OX and OY), the plane OXZ (defined by the axes OX and OZ), and the plane OYZ (defined by the axes OY and OZ). The three planes divide the space into eight octants, denoted Octant I, Octant II, Octant III, Octant IV, Octant V, Octant VI, Octant VII and Octant VIII (see Fig 2). A voxel with center point coordinates (x, y, z) are defined in terms of the basis vectors u , v , w as $P= x.u + y.v + z.w$. The sign of

voxel's center point coordinate (briefly called voxel's coordinate) is depending on the Octant in which the point lays.

Octant I	(+, +, +)	Octant II	(-, +, +)
Octant III	(-, -, +)	Octant IV	(+, -, +)
Octant V	(+, +, -)	Octant VI	(-, +, -)
Octant VII	(-, -, -)	Octant VIII	(+, -, -)

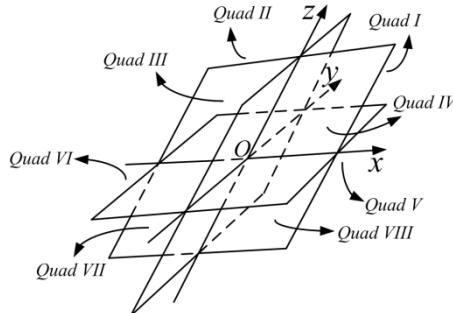


Fig. 2. The octants divided by axes Ox , Oy and Oz

A BCC grid line between voxels (x_1, y_1, z_1) and (x_2, y_2, z_2) is defined as an ordered sequence of voxels starting at coordinates (x_1, y_1, z_1) and finishing at coordinates (x_2, y_2, z_2) where a given voxel is included in the sequence a single time (no loops), and every voxel is connected through a face adjacency to the next voxel in the sequence.

3 Adjunct Parallelepiped Space

To extent the Bresenham algorithm in 3D to BCC grid, a sheared voxelized space called the adjunct parallelepiped space is needed to trace out the major voxels intersected, then remaps to the Voronoi cells of the samples on the BCC grid. We will explain how to build the adjunct parallelepiped space in this section.

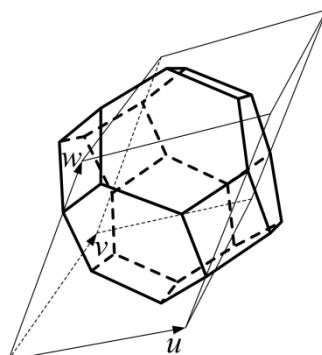


Fig. 3. Voxel $h(x, y, z)$ and its adjunct parallelepiped $p(x, y, z)$

With the voxelized space based on the BCC grid \mathbf{N} defined in section 2, one can define an auxiliary space, called the adjunct parallelepiped space and denoted \mathbf{N}^1 . Each voxel $h(x, y, z)$ in the space \mathbf{N} is associate to a parallelepiped $p(x, y, z)$ in the space \mathbf{N}^1 , called adjunct to $h(x, y, z)$, with the same center and with sides determined by the basis vectors of the space \mathbf{N} , i.e., \mathbf{u} , \mathbf{v} and \mathbf{w} (see Fig 3). This correspondence defines a parallelepiped discretized space \mathbf{N}^1 adjunct to \mathbf{N} with the same center and basis vectors.

Now considering more in detail the relation between the spaces \mathbf{N} and \mathbf{N}^1 , the adjacency of their cells is studied. We label a voxel from \mathbf{N} (resp. a parallelepiped from \mathbf{N}^1) by the coordinates of its center. Let $h(i, j, k)$ be a voxel from \mathbf{N} defined in terms of the basis vectors \mathbf{u} , \mathbf{v} , \mathbf{w} and $p(i, j, k)$ the corresponding parallelepiped from \mathbf{N}^1 . Now we get the following neighbor relations of parallelepipeds and the corresponding voxels shown in table 1. There are 13 adjacencies listed which correspond to half of the 26 possible adjacencies in a cubic-like space such as the adjunct parallelepiped space when considering face, edge and vertex adjacencies. The adjacencies can be reduced to 13 due to the reflection symmetries. The 26 adjacencies of the parallelepiped are more than the 14 possible adjacencies of the truncated octahedron and that that is the reason why 12 of the parallelepiped adjacencies lead to pairs of disjoint BCC voxels. Again, due to reflection symmetries, the table lists only six out of these 12 cases.

Table 1. Neighbor relations of parallelepipeds in space \mathbf{N}^1 and adjacencies of their corresponding voxels in space \mathbf{N}

Form of the pairs	Parallelepipeds neighbor relation	Voxels neighbor relation
(i, j, k) and $(i+1, j, k)$	Face-adjacency	Face-adjacency
(i, j, k) and $(i, j+1, k)$	Face-adjacency	Face-adjacency
(i, j, k) and $(i, j, k+1)$	Face-adjacency	Face-adjacency
(i, j, k) and $(i+1, j, k-1)$	Edge-adjacency	Face-adjacency
(i, j, k) and $(i+1, j-1, k)$	Edge-adjacency	Face-adjacency
(i, j, k) and $(i, j+1, k-1)$	Edge-adjacency	Face-adjacency
(i, j, k) and $(i, j+1, k+1)$	Edge-adjacency	Disjoint
(i, j, k) and $(i+1, j, k+1)$	Edge-adjacency	Disjoint
(i, j, k) and $(i+1, j+1, k)$	Edge-adjacency	Disjoint
(i, j, k) and $(i+1, j+1, k-1)$	Vertex-adjacency	Face-adjacency
(i, j, k) and $(i+1, j+1, k+1)$	Vertex-adjacency	Disjoint
(i, j, k) and $(i+1, j-1, k+1)$	Vertex-adjacency	Disjoint
(i, j, k) and $(i-1, j+1, k+1)$	Vertex-adjacency	Disjoint

4 Voxelization Algorithm for 3D BCC Line

The algorithm is similar to the Bresenham algorithm, which has been originally designed for the cubic grid. The key idea is to perform the line-drawing in the

parallelepiped space exploiting the one-to-one correspondence between the parallelepipeds and the truncated octahedron voxels of the BCC grid. The obtained line is sometimes disconnected, therefore additional BCC voxels are identified which guarantee the face-adjacency and fit onto the line best.

Let the two endpoints of the line to be voxelized on the BCC grid have the BCC coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2) . Then (x_1, y_1, z_1) and (x_2, y_2, z_2) are also the corresponding parallelepiped coordinates of the two endpoints. Let $\Delta x = |x_2 - x_1|$, $\Delta y = |y_2 - y_1|$ and $\Delta z = |z_2 - z_1|$, $xsign = SIGN(x_2 - x_1)$, $ysign = SIGN(y_2 - y_1)$ and $ysign = SIGN(z_2 - z_1)$. To draw a BCC grid line between (x_1, y_1, z_1) and (x_2, y_2, z_2) . The algorithm solves the problem by the following steps:

Step 1. Select the octant in which the line lays as introduced in section 2. Initialize current parallelepiped and corresponding voxel at the point (x_1, y_1, z_1) .

Step 2. Determine the next parallelepiped in the adjunct parallelepiped space.

Step 3. Determine the next BCC voxels with the help of the one-to-one correspondence between the parallelepipeds and the BCC voxels.

Step 4. If voxel (x_2, y_2, z_2) has not yet been reached, go to step 2. Otherwise, the procedure terminates.

5 Analyses and Comparisons of Algorithms

The proposed algorithm uses only integer arithmetic, and only addition, subtraction, shifting, and comparison are employed except uses two multiplications in the decision for the two voxels to ajoin the line. At the same time, with the decision, three voxels are generated in only one step.

In order to prove the efficiency of the new algorithm, it is compared to the existing line tracing algorithm on BCC grid: Ibáñez-Hamitouche-Roux's algorithm (IHR) [19]. The operations at each step in the Ibáñez-Hamitouche-Roux's algorithm, as published in their paper, are two vector addition and selecting the index of the minimum element in a vector. The number of numerical operations in each step is counted and shown in table 2. Here as the operations for updating the X , Y and Z coordinates of voxels being pierced are the same for both algorithms, they are omitted during the counting. As can be seen from table 2, the number of the new algorithm's arithmetic operations is smaller than the Ibáñez-Hamitouche-Roux's algorithm.

Table 2. Number of arithmetic operations at different line voxelization algorithms (number of operations per voxels is given in bracket)

Operation name	IHR algorithm	New algorithm generated voxel in step	New algorithm 1 generated voxels in step	New algorithm 2 generated voxels in step	New algorithm 3 generated voxels in step
Division	0	0	0	0	0
Multiple	0	0	0	2(0.67)	
Increase/decrease	7	3	3(1.5)	4(1.33)	
Comparison	3	3	5(2.5)	6(2)	

Both algorithms tested were implemented in C++ and executed on Intel's Pentium 4 2.53GHz. For testing, a parallelepiped was embedded in the grid of voxels. The line was voxelized by connecting the center of the parallelepiped and the points on the parallelepiped. By changing the edge length of the parallelepiped, different line segments were voxelized. The comparison of CPU time for both tested algorithms is shown in table 3. The times include the execution-time in the main loop and the execution-time needed to set the initial values outside the main loop. All input/output operations are excluded. As shown in table 3, the proposed algorithm is more efficient than the Ibáñez-Hamitouche-Roux's algorithm.

Table 3. CPU times (seconds) of line voxelization

Length of parallelepiped's edge	IHR algorithm	New algorithm
50	3.0315	2.6255
60	6.8135	4.4065
70	10.5475	7.0315
80	14.6415	10.3950
90	18.2550	14.9375
100	26.9535	20.3135

6 Conclusion

A voxelization algorithm for a given 3D continuous line generates a minimal set of connected BCC voxels that approximates a straight line segment as best possible. The basic criterion of our algorithm for tracing the discrete approximation of the line is to find at each step the BCC voxel that is closest to the line among all the possible voxels allowed by the connectivity of the grid. The algorithm is based on the 3D Bresenham line-drawing algorithm and the adjunct parallelepiped space. The algorithm is explained in detail and implementation of the algorithm is as similar as implementation of 3D Bresenham algorithm. Up to three voxels are chosen in only one step.

Acknowledgments. The authors thank Dr. Luis Ibáñez for the code of literature [19]. This work was supported by the Research Project (L2010096) of Science and Technology of Education Department of Liaoning Province and the Fundamental Research Funds for the Central Universities.

References

1. Žalík, B., Clapworthy, G., Oblonšek, Č.: An Efficient Code-Based Voxel-Traversing Algorithm. Computer Graphics Forum 16, 119–128 (1997)
2. Liu, Y.K., Žalík, B., Yang, H.: An Integer One-Pass Algorithm for Voxel Traversal. Computer Graphics Forum 23, 167–172 (2004)
3. Bresenham, J.E.: Algorithm for Computer Control of a Digital Plotter. IBM System Journal 4, 25–30 (1965)

4. Liu, X.W., Cheng, K.: Three-dimensional extension of Bresenham's algorithm and its application in straight-line interpolation. *Journal of Engineering Manufacture* 216, 459–463 (2002)
5. Kaufman, A., Shimony, E.: 3D scan-conversion algorithms for voxel-based graphics. In: *Proceedings of the 1986 Workshop on Interactive 3D Graphics*, pp. 45–75 (1986)
6. Cohen-Or, D., Kaufman, A.: 3D Line Voxelization and Connectivity Control. *IEEE Computer Graphics and Applications* 17, 80–87 (1997)
7. Toutant, J.L.: Characterization of the Closest Discrete Approximation of a Line in the 3-Dimensional Space. In: *International Symposium on Visual Computing* (2006)
8. Buzer, L.: A linear incremental algorithm for naive and standard digital lines and planes recognition. *Graphical Models* 65, 61–76 (2003)
9. Burns, M., Klawe, J., Rusinkiewicz, S., Finkelstein, A., DeCarlo, D.: Line drawings from volume data. In: *ACM SIGGRAPH 2005*, pp. 512–518 (2005)
10. Brimkov, V.E., Barneva, R.P.: Analytical Honeycomb Geometry for Raster and Volume Graphics. *The Computer Journal* 48, 180–199 (2005)
11. Bernhard, F., Alireza, E., Dimitri, V.D.V., Torsten, M.: Efficient volume rendering on the body centered cubic lattice using box splines. *Computers & Graphics* 34(4), 409–423 (2010)
12. Csebfalvi, B.: An Evaluation of Prefiltered B-Spline Reconstruction for Quasi-Interpolation on the Body-Centred Cubic Lattice. *IEEE Transactions on Visualization and Computer Graphics* 16(3), 499–512 (2010)
13. He, L.J., Liu, Y.K., Sun, S.C.: A line generation algorithm on 3D face-centered cubic grid. *Chinese Journal of Computers* 33(12), 2407–2416 (2010)
14. He, L.J., Liu, Y.K., Bao, S.Z.: 3D Line generation algorithm on hexagonal prism grids. In: *2008 International Congress on Image and Signal Processing*, pp. 741–745 (2008)
15. Miller, E.G.: Alternative tilings for improved surface area estimates by local counting algorithms. *Computer Vision and Image Understanding* 74, 193–211 (1999)
16. Strand, R.: Surface skeletons in grids with non-cubic voxels. In: *Pattern Recognition 2004*, pp. 548–551 (2004)
17. Carr, H., Theußl, T., Möller, T.: Isosurfaces on Optimal Regular Samples. In: *Proceedings of Eurographics Visualization Symposium 2003*, pp. 39–48 (2003)
18. Theusl, T., Moller, T., Groller, M.E.: Optimal regular volume sampling. In: *Proceedings of the Conference on Visualization 2001*, pp. 91–98 (2001)
19. Ibáñez, L., Hamitouche, C., Roux, C.: A Vectorial Algorithm for Tracing Discrete Straight Lines in N-Dimensional Generalized Grids. *IEEE Transactions on Visualization and Computer Graphics* 7, 97–108 (2001)

Jackson's Theorem in Hardy-Sobolev Type Spaces in the Unit Polydiscs

Yingwei Chen¹, Zhijun Wang¹, and Wenlei Dong²

¹ College of Mathematics and Statistics,
Hebei University of Economics and Business,
Shijiazhuang, 050061 Hebei, China

² Department of Basic Courses,
Shijiazhuang Institute of Railway Technology,
Shijiazhuang, 050041 Hebei, China
cyingwei@tom.com

Abstract. The purpose of this article is to establish Jackson type inequality in the unit polydiscs U^n of \mathbb{C}^n for Hardy-Sobolev type spaces $F_\alpha^{p,q}(U^n)$. Namely,

$$E_{\vec{k}}(f, F_\alpha^{p,q}(U^n)) \lesssim \omega_r(\vec{1/k}, f, F_\alpha^{p,q}(U^n)),$$

where $E_{\vec{k}}(f, F_\alpha^{p,q}(U^n))$ is the deviation of the best approximation of $f \in F_\alpha^{p,q}(U^n)$ by polynomials of degree at most k_j about the j -th variable z_j with the corresponding moduli of smoothness.

Keywords: Hardy-Sobolev type spaces, Jackson's theorem, Polynomial approximation.

1 Introduction

Jackson's theorem (see [4,10]) is an important result in the theory of approximation treating the deviation of the best approximation of a function by polynomials. It has been established for various classes of functions ([1,2,4,7,8]).

In this article the best approximation by polynomials with vector degrees is considered. More precisely, for any semi-normed space \mathcal{X} on the complex unit polydiscs U^n with semi-norm $\|\cdot\|_{\mathcal{X}}$, we define the best approximation of f of order \vec{k} by \mathcal{X} as

$$E_{\vec{k}}(f, \mathcal{X}) \triangleq \inf_{Q_k \in \mathcal{W}_{\vec{k}}} \|f - Q_k\|_{\mathcal{X}},$$

where $\mathcal{W}_{\vec{k}} \triangleq \mathcal{W}_{\vec{k}}(U^n)$ is the polynomial space of vector degrees, i.e., the set of all polynomials whose degrees of the j -th variables are at most k_j for $j = 1, \dots, n$. Here the degree \vec{k} is a vector $\vec{k} = (k_1, \dots, k_n)$.

For any $f \in \mathcal{X}$, and $\vec{\delta} = (\delta_1, \dots, \delta_n) > 0$, we define the moduli of smoothness of f as

$$\omega_r(\vec{\delta}, f, \mathcal{X}) \triangleq \sup_{|h| \in I_{\delta,n}} \|\Delta_h^r f(z)\|_{\mathcal{X}},$$

where $|h| = (|h_1|, \dots, |h_n|)$, $I_{\delta,n} = [0, \delta_1) \times \dots \times [0, \delta_n)$ and

$$\Delta_h^r f(z) = \sum_{m=0}^r (-1)^{r-m} \binom{r}{m} f(ze^{imh}).$$

For any nonnegative function f and g , we denote $f \lesssim g$, if there exists a positive constant C such that $f \leq Cg$. We also write $f \sim g$ if $f \lesssim g$ and $g \lesssim f$.

Recently, Ren and Wang [9] considered an extension of Jackson's theorem to some holomorphic function spaces in U^n , such as Hardy spaces $H^p(U^n)$.

Theorem 1.1. [9] Let $0 < p \leq +\infty$, $f \in H^p(U^n)$, then for any $k \in \mathbb{N}^n$,

$$E_{\vec{k}}(f, H^p(U^n)) \lesssim \omega(\overrightarrow{1/k}, f, H^p(U^n)).$$

Let $H(U^n)$ be the set of all holomorphic functions in U^n with Shilov boundary T^n (see [11]). In polar coordinates we write $z = R\zeta$ with $R \in I^n \triangleq [0, 1]^n$ and $\zeta \in T^n$. The Hardy-Sobolev type space $F_\alpha^{p,q}(U^n)$ (see [12]) is defined to be the set of all holomorphic functions $f \in H(U^n)$ with

$$\|f(z)\|_{F_\alpha^{p,q}(U^n)} = \left(\int_{T^n} \left(\int_{I^n} |f(R\zeta)|^q (1-R)^\alpha R dR \right)^{\frac{p}{q}} dm_n(\zeta) \right)^{1/p} < +\infty,$$

where $\alpha > -1$, $dR = dR_1 \cdots dR_n$, $Rz = (R_1 z_1, \dots, R_n z_n)$, $|z_1 - z_2|^\gamma = \prod_{j=1}^n |z_1^j - z_2^j|^\gamma$, $\gamma \in R$, $z_1, z_2 \in \mathbb{C}^n$, $z_m = (z_m^1, \dots, z_m^n)$, $m = 1, 2$, $z_1^j \neq z_2^j$.

It is easy to see that (see [5]) $F_{2\alpha-1}^{p,2} = H_\alpha^p(U)$, $0 < p < \infty$, and $F_\alpha^{p,p}(U^n) = A_\alpha^p(U^n)$, where $H_\alpha^p(U)$ and $A_\alpha^p(U^n)$ are the known Hardy-Sobolev and Bergman-Djrbashian class in the unit disk U and in the unit polydiscs U^n , respectively.

The main purpose of this article is not only to extend Jackson's theorems to Hardy-Sobolev type spaces in the unit polydiscs, but also to obtain some results by moduli of smoothness of higher order. Through the same approach, Theorem 1.1 also can be easily extended to the results of higher order.

2 Preliminaries

For any $z = (z_1, \dots, z_n) \in \mathbb{C}^n$, we denote

$$\|z\|_{\max} = \max_{j=1, \dots, n} |z_j|,$$

then U^n is the unit ball with respect to the norm above.

We shall use the following notations. We write

$$\vec{k}^m = \prod_{j=1}^n k_j^m, \quad \overrightarrow{1/k} = \left(\frac{1}{k_1}, \dots, \frac{1}{k_n} \right), \quad \vec{\lambda}\delta = (\lambda_1\delta_1, \dots, \lambda_n\delta_n)$$

for any $k = (k_1, \dots, k_n) \in \mathbb{N}^n$, $\lambda = (\lambda_1, \dots, \lambda_n)$, $\delta = (\delta_1, \dots, \delta_n)$, and non-negative integer m . We also write $d\varphi = d\varphi_1 \cdots d\varphi_n$. For any

multi-index $\alpha = (\alpha_1, \dots, \alpha_n)$, we denote $|\alpha| = \alpha_1 + \dots + \alpha_n$, $z^\alpha = z_1^{\alpha_1} \cdots z_n^{\alpha_n}$, and by $\alpha > 0$ we mean that each component $\alpha_j > 0$.

Our approach is to construct the best approximation of polynomials with a kind of complex measures on T^n , whose total variations are given by generalized Jackson's kernels:

$$T_k^\beta(\theta) \triangleq \left(\frac{\sin \frac{k\theta}{2}}{\sin \frac{\theta}{2}} \right)^{2\beta}.$$

Lemma 2.1. (See [4]) For each $k, \beta \in \mathbb{N}$,

$$B_{k,\beta} \triangleq \int_{-\pi}^{\pi} T_k^\beta(\theta) d\theta \sim k^{2\beta-1}, \quad k \rightarrow +\infty$$

and there is a constant C_β such that

$$\frac{1}{B_{k,\beta}} \int_0^\pi \theta^u T_k^\beta(\theta) d\theta \leq C_\beta k^{-u}, \quad u = 0, 1, \dots, 2\beta - 2.$$

Now we introduce the required normalized complex measure $d\mu_k^\rho(\varphi)$ on $[-\pi, \pi]^n$. We fix an $a > 0$ and define

$$d\mu_k^\rho(\varphi) = i^n C_n^{\vec{k}} \prod_{j=1}^n (\rho_j e^{i\varphi_j})^{1-k_j} \left(\frac{1 - (\rho_j e^{i\varphi_j})^{k_j+1}}{1 - \rho_j e^{i\varphi_j}} \right)^{a+1} d\varphi,$$

where

$$C_n^{\vec{k}} = \prod_{j=1}^n (2\pi i A_{k_j}^a)^{-1}, \quad A_{k_j}^a = \frac{\Gamma(k_j + a)}{\Gamma(k_j)\Gamma(a+1)}.$$

Then $d\mu_k^\rho(\varphi)$ are normalized measures on $[-\pi, \pi]^n$ for any $\rho \in (0, 1]^n$ and $k \in \mathbb{N}^n$.

There are two important facts related to these measures. One is the total variation of $dv_k \triangleq d\mu_k^\rho(\varphi)$ for $\rho = (1, \dots, 1)$ given by the generalized Jackson kernels. In fact, by direct calculation we have

$$d|v_k| = |C_n^{\vec{k}}| \prod_{j=1}^n T_{k_j}^{\frac{a+1}{2}}(\varphi_j) d\varphi.$$

For simplicity, write $a(r, m) = (-1)^{-m-1} \binom{r}{m}$. The other one is the associated operators P_k :

$$P_k[f](z) = \sum_{m=1}^r a(r, m) \int_{[-\pi, \pi]^n} f((\rho e^{i\varphi})^m z) d\mu_k^\rho(\varphi), \quad (1)$$

which will provide the best approximation of polynomials.

Lemma 2.2. (See [9]) Let f be holomorphic on U^n and $\rho \in (0, 1]^n$.

- (A) $P_k[f](z)$ is a polynomial of degree at most $|k| - n$; moreover its j -th variable's degree is at most $k_j - 1$, $j = 1, \dots, n$; $P_k[1] \equiv 1$.
- (B) $P_k[f](z)$ has another integral formula: for any $\rho \in (0, 1)^n$,

$$\sum_{m=1}^r a(r, m) C_n^{\vec{k}} \int_{\rho T^n} f(\lambda^m z) \prod_{j=1}^n \frac{\lambda_j^{-k_j}}{(1 - \lambda_j)^{(a+1)}} d\lambda. \quad (2)$$

3 Hardy-Sobolev Type Spaces

In this section, we consider the polynomial approximation in Hardy-Sobolev type spaces.

If we consider the measure on $[-\pi, \pi)^n$:

$$d\nu_k^{\rho, \eta, a}(\varphi) \triangleq |C_n^{\vec{k}}|^{\eta} \prod_{j=1}^n \frac{\rho_j^{\eta(1-k_j)}}{(1 - \rho_j)^{1-\eta}} \prod_{j=1}^n T_{k_j+1}^{\frac{\eta(a+1)}{2}}(\varphi_j) d\varphi$$

for any $0 < \rho_j < 1$, $\eta > 0$, and $a > 0$, we have the following estimates.

Lemma 3.1. For any $0 < \eta \leq 1$, $k = (k_1, \dots, k_n) \in \mathbb{N}^n$, and $f \in H(U^n)$,

$$|f(z) - P_k[f](z)|^\eta \lesssim \int_{[-\pi, \pi)^n} |\Delta_\varphi^r f(z)|^\eta d\nu_k^{\rho, \eta, a}(\varphi). \quad (3)$$

Proof: It is well known that [6] if g is holomorphic in the closed unit disc \overline{U} , $0 < p \leq 1$, $0 < r < 1$, then

$$\left(\int_{-\pi}^{\pi} |g(re^{i\theta})| d\theta \right)^p \lesssim (1 - r)^{p-1} \int_{-\pi}^{\pi} |g(e^{i\theta})|^p d\theta. \quad (4)$$

By Lemma 2.2, we find that for any $\rho \in (0, 1)^n$

$$f(z) - P_k[f](z) = (-1)^{-r} \int_{[-\pi, \pi)^n} \Delta_\varphi^r f(z) d\mu_k^\rho(\varphi), \quad (5)$$

where

$$d\mu_k^\rho(\varphi) = i^n C_n^{\vec{k}} \prod_{j=1}^n (\rho_j e^{i\varphi_j})^{1-k_j} \left(\frac{1 - (\rho_j e^{i\varphi_j})^{k_j+1}}{1 - \rho_j e^{i\varphi_j}} \right)^{a+1} d\varphi.$$

Disregarding the constant, $d\mu_{k,n}^\rho(\varphi)$ equals

$$(\rho_n e^{i\varphi_n})^{1-k_n} \left(\frac{1 - (\rho_n e^{i\varphi_n})^{k_n+1}}{1 - \rho_n e^{i\varphi_n}} \right)^{a+1} d\mu_{k,n-1}^{\rho'}(\varphi') d\varphi_n,$$

where $\rho = (\rho', \rho_n)$ and $\varphi = (\varphi', \varphi_n)$.

To deal with the integral over $[-\pi, \pi]^n$ in (5), we first rewrite it as an iterated integral and then consider it as the integral over $[-\pi, \pi]$. Therefore,

$$|f(z) - P_k[f](z)| = \int_{[-\pi, \pi)} |g(\rho e^{i\varphi}, z)| d\varphi_n,$$

where $|g(\rho e^{i\varphi}, z)|$ equals

$$\left| \int_{[-\pi, \pi)^{n-1}} \Delta_\varphi^r f(z) d\mu_{k,n-1}^{\rho'}(\varphi') |\rho_n^{1-k_n} \frac{1 - (\rho_n e^{i\varphi_n})^{k_n+1}}{1 - \rho_n e^{i\varphi_n}}|^{a+1} \right|.$$

Let $h(\rho e^{i\varphi}, z)$ denote the integral

$$\int_{[-\pi, \pi)^{n-1}} \Delta_\varphi^r f(z) \left(\frac{1 - (\rho_n e^{i\varphi_n})^{k_n+1}}{1 - \rho_n e^{i\varphi_n}} \right)^{a+1} d\mu_{k,n-1}^{\rho'}(\varphi').$$

Then $g(\rho e^{i\varphi}, z) = \rho_n^{1-k_n} |h(\rho e^{i\varphi}, z)|$, so

$$|f(z) - P_k[f](z)| \lesssim \rho_n^{1-k_n} \int_{[-\pi, \pi)} |h(\rho e^{i\varphi}, z)| d\varphi_n.$$

Since $h(\rho e^{i\varphi}, z)$ is a holomorphic function of $\lambda_n = \rho_n e^{i\varphi_n}$ in $A(U)$ for any fixed $z \in U^n$, we can invoke inequality (4). The resulting integral over $[-\pi, \pi)^{n-1}$ can be dealt with the same procedure as above. Finally, we can deduce that

$$|f(z) - P_k[f](z)|^\eta \lesssim \int_{[-\pi, \pi)^n} |\Delta_\varphi^r f(z)|^\eta d\nu_k^{\rho, \eta, a}(\varphi),$$

as desired.

Lemma 3.2. Assume $0 < \eta \leq 1$ and $\eta(a+1) \geq r+2$. Then for any $k \in \mathbb{N}^n$ and $\rho_j = 1 - \frac{1}{k_j}$, ($j = 1, \dots, n$),

$$\int_{[-\pi, \pi)^n} (\|k\varphi\|_{\max} + 1)^r d\nu_k^{\rho, \eta, a}(\varphi) = O(1), \quad \text{as } m \rightarrow \infty.$$

Proof: By Lemma 2.1, the proof is similar for high-order case to that of Lemma 2.2 in [9].

Lemma 3.3. Let $0 < p, q \leq \infty$, $s = \min\{1, p, q\}$, $f \in F_\alpha^{p,q}(U^n)$, then

$$\omega_r^s(\overrightarrow{\lambda\delta}, f, F_\alpha^{p,q}(U^n)) \leq (\|\lambda\|_{\max} + 1)^r \omega_r^s(\overrightarrow{\delta}, f, F_\alpha^{p,q}(U^n)).$$

Proof: Let $m = [\|\lambda\|_{\max}]$ denote the greatest integer less than or equal to $\|\lambda\|_{\max}$. Just like in [10], by induction on r , we easily have

$$\Delta_{mh}^r f(z) = \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \Delta_h^r f(ze^{\sum_{j=1}^r k_j h}). \quad (6)$$

Denoting $s = \min\{1, p, q\}$. We claim that

$$\begin{aligned} & \|\Delta_{mh}^r f(z)\|_{F_\alpha^{p,q}(U^n)}^s \\ & \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \|\Delta_h^r f(z e^{\sum_{j=1}^r k_j h})\|_{F_\alpha^{p,q}(U^n)}^s \\ & = m^r \|\Delta_h^r f(z)\|_{F_\alpha^{p,q}(U^n)}^s. \end{aligned} \quad (7)$$

With this claim, by the monotone property of moduli of smoothness and (6), we obtain

$$\begin{aligned} \omega_r^s(\overrightarrow{\lambda\delta}, f, F_\alpha^{p,q}(U^n)) & \leq \omega_r^s((m+1)\overrightarrow{\delta}, f, F_\alpha^{p,q}(U^n)) \\ & \leq (m+1)^r \omega_r^s(\overrightarrow{\delta}, f, F_\alpha^{p,q}(U^n)) \\ & \leq (||\lambda||_{max} + 1)^r \omega_r^s(\overrightarrow{\delta}, f, F_\alpha^{p,q}(U^n)). \end{aligned}$$

It remains to prove claim (7). For simplicity, let

$$M_q(\zeta, f(z)) = \left(\int_{I_n} |f(R\zeta)|^q (1-R) R dR \right)^{\frac{1}{q}},$$

where $z = R\zeta$, $\zeta \in T^n$. When $1 \leq q < +\infty$, it follows from (6) and Minkowski's inequality that

$$M_q(\zeta, \Delta_{mh}^r f(z)) \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} M_q(\zeta, \Delta_h^r f(z e^{\sum_{j=1}^r k_j h})). \quad (8)$$

If $0 < q < 1$, we have

$$M_q^q(\zeta, \Delta_{mh}^r f(z)) \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} M_q^q(\zeta, \Delta_h^r f(z e^{\sum_{j=1}^r k_j h})), \quad (9)$$

since $(|a| + |b|)^q \leq |a|^q + |b|^q$ for any $(a, b) \in \mathbb{C}^2$.

We split the discussion into four cases.

i) $1 \leq p < +\infty$, $1 \leq q < +\infty$.

From Minkowski's inequality and (8), we get

$$\|\Delta_{mh}^r f(z)\|_{F_\alpha^{p,q}(U^n)} \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \|\Delta_h^r f(z e^{\sum_{j=1}^r k_j h})\|_{F_\alpha^{p,q}(U^n)}.$$

ii) $0 < p < 1$, $1 \leq q < +\infty$.

Since $p/q \leq 1$, by (8) we have

$$\begin{aligned} \|\Delta_{mh}^r f(z)\|_{F_\alpha^{p,q}(U^n)}^p & \leq \int_{T^n} \left(\sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} M_q^q(\zeta, \Delta_h^r f(z e^{\sum_{j=1}^r k_j h})) \right)^{\frac{p}{q}} dm_n(\zeta) \\ & \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \left\| \Delta_h^r f(z e^{\sum_{j=1}^r k_j h}) \right\|_{F_\alpha^{p,q}(U^n)}^p. \end{aligned}$$

iii) $0 < q < 1, 0 < p \leq q \leq 1$

Since $p/q \leq 1$, we have $(|a| + |b|)^{p/q} \leq |a|^{p/q} + |b|^{p/q}$, so that (9) implies

$$\begin{aligned} \|\Delta_{mh}^r f(z)\|_{F_\alpha^{p,q}(U^n)}^p &\leq \int_{T^n} \left(\sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} M_q^q(\zeta, \Delta_h^r f(z e^{\sum_{j=1}^r k_j h})) \right)^{\frac{p}{q}} dm_n(\zeta) \\ &\leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \|\Delta_h^r f(z e^{\sum_{j=1}^r k_j h})\|_{F_\alpha^{p,q}(U^n)}^p \end{aligned}$$

iv) $0 < q < 1, 0 < q \leq p < +\infty$

Since $\frac{p}{q} \geq 1$, from (9) and Minkowski's inequality,

$$\|\Delta_{mh}^r f(z)\|_{F_\alpha^{p,q}(U^n)}^q \leq \sum_{k_1=0}^{m-1} \cdots \sum_{k_r=0}^{m-1} \|\Delta_h^r f(z e^{\sum_{j=1}^r k_j h})\|_{F_\alpha^{p,q}(U^n)}^q.$$

With the above preparation we can establish the main result of this article.

Theorem 3.4. Let $0 < p, q < +\infty$, $f \in F_\alpha^{p,q}(U^n)$, then for any $k = (k_1, \dots, k_n) \in \mathbb{N}^n$,

$$E_{\vec{k}}(f, F_\alpha^{p,q}(U^n)) \lesssim \omega_r(\overrightarrow{1/k}, f, F_\alpha^{p,q}(U^n)).$$

Proof: Let $P_k[f]$ be the polynomial in (1) and let $s = \min\{1, p, q\}$. We claim that

$$\|f - P_k[f]\|_{F_\alpha^{p,q}(U^n)}^s \lesssim \int_{[-\pi, \pi)^n} \|\Delta_\varphi^r f(z)\|_{F_\alpha^{p,q}(U^n)}^s d\nu_k^{\rho, s, a}(\varphi). \quad (10)$$

Set $\eta = s$ in (3), we have

$$|f(z) - P_k[f](z)|^s \lesssim \int_{[-\pi, \pi)^n} |\Delta_\varphi^r f(z)|^s d\nu_k^{\rho, s, a}(\varphi).$$

Taking the q/s -th power, we have

$$|f(z) - P_k[f](z)|^q \lesssim \left(\int_{[-\pi, \pi)^n} |\Delta_\varphi^r f(z)|^s d\nu_k^{\rho, s, a}(\varphi) \right)^{\frac{q}{s}}.$$

Then integrate over I^n and apply Minkowski's inequality to yield

$$M_q^s(\zeta, f - P_k[f]) \lesssim \int_{[-\pi, \pi)^n} M_q^s(\zeta, \Delta_\varphi^r f(z)) d\nu_k^{\rho, s, a}(\varphi).$$

The claim then follows from Minkowski's inequality.

From (10), we have

$$\|f - P_k[f]\|_{F_\alpha^{p,q}(U^n)}^s \lesssim \int_{[-\pi, \pi)^n} \omega_r(\vec{\varphi}, f, F_\alpha^{p,q}(U^n)) d\nu_k^{\rho, s, a}(\varphi),$$

where $\vec{\varphi}$ denotes the vector $(|\varphi_1|, \dots, |\varphi_n|)$.

Pick $a > 0$ such that $s(a + 1) \geq r + 2$. By Lemmas 3.3 and 3.2, we obtain

$$\begin{aligned} & \|f - P_k[f]\|_{F_\alpha^{p,q}(U^n)}^s \\ & \lesssim \omega_r^s \left(\overrightarrow{1/k}, f, F_\alpha^{p,q}(U^n) \right) \int_{[-\pi,\pi)^n} (\|k\varphi\|_{\max} + 1)^r d\nu_k^{\rho,s,a}(\varphi) \\ & \lesssim \omega_r^s \left(\overrightarrow{1/k}, f, F_\alpha^{p,q}(U^n) \right). \end{aligned}$$

We finally get

$$E_{\overrightarrow{k}}(f, F_\alpha^{p,q}(U^n)) \lesssim \omega_r \left(\overrightarrow{1/k}, f, F_\alpha^{p,q}(U^n) \right).$$

Acknowledgment. This work is supported by National Natural Science Foundation of China (Grant No. 11126246).

References

1. Bhaya, E.S.: On Jackson's Theorem. *Int. J. Contemp. Math. Sciences* 7, 1149–1154 (2012)
2. Chen, Y.W., Ren, G.B.: Jackson's Theorem in Q_p Spaces. *Science in China Series A: Mathematics* 53, 367–372 (2010)
3. Chen, Y.W., Liu, Y.J.: Strong Converse Inequality in Hardy Type Spaces \mathcal{A}_μ . *J. Hebei Norm. Univ. Nat. Sci. Ed.* 35, 336–342 (2011) (Chinese)
4. Devore, R.A., Lorentz, G.G.: *Constructive Approximation*. Springer, Berlin (1993)
5. Fefferman, C., Stein, E.: H^p Spaces of Several Variables. *Acta Math.* 192, 137–173 (1972)
6. Garnett, J.B.: *Bounded Analytic Functions*. Academic Press, New York (1981)
7. Ivanov, A.V., Ivanov, V.I.: Jackson's Theorem in the Space $L_2(\mathbb{R}^d)$ with Power Weight. *Mathematical Notes* 88, 140–143 (2010)
8. Ren, G.B., Chen, Y.W.: Gradient Estimates and Jackson's Theorem in Q_μ Spaces Related to Measures. *J. Approx. Theory* 155, 97–110 (2008)
9. Ren, G.B., Wang, M.Z.: Holomorphic Jackson's Theorems in Polydiscs. *J. Approx. Theory* 134, 175–198 (2005)
10. Lorentz, G.G.: *Approximation of Function*. Athena Series. Holt, Rinehart and Winston, New York (1966)
11. Rudin, W.: *Function Theory in Polydiscs*. Benjamin, New York (1969)
12. Shamoyan, R.F.: On the Representation of Linear Continuous Functionals in Spaces of Analytic Functions of the Hardy-Sobolev Type in a Polydisk. *Ukrainian Math. J.* 55, 812–831 (2003)

On the Upper Bounds of Local Inverse Signed Edge Domination Numbers in Graphs

Zhifang Feng, Wensheng Li, Huaming Xing, and Qiongyu Ma

Dept. of Math. and Science, Langfang Teachers College, 065000, Langfang, China
nacyfzf@yahoo.com.cn, hmxing@tust.edu.cn

Abstract. In this paper we introduce the concept of local inverse signed edge domination in graphs. For a graph G , the concepts of local inverse signed edge domination function(LRSEDF) and local inverse signed edge domination number(LRSEDN) are gived. At the same time, we obtain four upper bounds of LRSEDN for general graphs G .

Keywords: Local Signed Edge Domination Function, Local Inverse Signed Edge Domination Function, Local Inverse Signed Edge Domination Number.

1 Introduction

In recent years, the concept of domination in graphs, with its many variations, now is well studied in graph theory. Such as signed domination[1], minus domination[2]. Nowadays, domination research does not only focus on vertices[3], studing on edges is continuously enriched. Signed edge domination[4] and local signed edge domination[5] are defined by B. G. Xu in 2001 and 2007, respectively. By then, some research on that is given by him[6-12]. we introduce the new concept of local inverse signed edge domination of a graph G .

2 Definition

Let $G = (V, E)$ be a graph. If $v \in V$, then $N_G(v) = \{u | uv \in E\}$ is called the open neighbourhood of the vertex v , and $N_G[v] = N_G(v) \cup \{v\}$ is the closed one. The degree of v is denote by $d_G(v) = |N(v)|$, $\Delta(G)$ and $\delta(G)$ are the maximun one and minimun one, respectively. $d_G(v)$, $\Delta(G)$ and $\delta(G)$ is denote by $d(v)$, Δ and δ , for convenience. $E_G(v)$ is the edge neighbourhood of vertex v , that is $E_G(v) = \{uv \in E | u \in V\}$. If $e = uv \in E$, then $N_G[e] = \{u'v' \in E | u' = u \text{ or } v' = v\}$ is called the closed edge-neighbourhood of e in G , and $N_G(e) = N_G[e] \setminus \{e\}$ is the open one. For simplicity, sometimes $N_G(v)$, $N_G[v]$, $E_G(v)$, $N_G(e)$, $N_G[e]$ are denoted by $N(v)$, $N[v]$, $E(v)$, $N(e)$, $N[e]$.

Definition 1. Let $G = (V, E)$ be graph without isolated vertices. A function $f : E \rightarrow \{-1, 1\}$ is called the local signed edge domination function(LSEDF) of G if $\sum_{e \in E(v)} f(e) \geq 1$ for every $v \in V(G)$. The local signed edge domination number of G is defined as $\gamma'_l(G) = \min\{\sum_{e \in E} f(e) \mid f \text{ is an LSEDF of } G\}$. Obviously, $|\gamma'_l(G)| \leq |E(G)|$.

Definition 2. Let $G = (V, E)$ be graph without isolated vertices. A function $f : E \rightarrow \{-1, 1\}$ is called a local inverse signed edge dominating function (LRSEDF) of G if $\sum_{e \in E(v)} f(e) \leq 0$ for every $v \in V(G)$. The local inverse signed edge domination number of G is defined as $\tilde{\gamma}'_l(G) = \max\{\sum_{e \in E} f(e) \mid f \text{ is an LRSEDF of } G\}$. A LRSEDF f is called a $\tilde{\gamma}'_l(G)$ -function if $w(f) = \tilde{\gamma}'_l(G)$.

Obviously, $|\tilde{\gamma}'_l(G)| \leq |E(G)|$. It seems naturally to define $\tilde{\gamma}'_l(\bar{K}_n) = 0$ for all totally disconnected graphs \bar{K}_n .

Definition 3. An LRSEDF f of G is called a maximal LRSEDF, if there is no any other LRSEDF g ($g \neq f$) such that $g > f$, for which $g(e) \geq f(e)$ for every $v \in V$.

Definition 4. f is called the maximum local inverse signed edge domination function, if the $w(f) = \max\{w(g) \mid g \text{ is an LRSEDF of } G\}$. f is called a $\tilde{\gamma}'_l(G)$ -function.

3 Main Results

Proposition 1. Let $G = (V, E)$ be a nonempty simple graph, $\tilde{\gamma}'_l(G) \equiv |E(G)| \pmod{2}$.

Proposition 2. For any two disjoint graphs G_1 and G_2 , $\tilde{\gamma}'_l(G_1 \cup G_2) = \tilde{\gamma}'_l(G_1) + \tilde{\gamma}'_l(G_2)$.

Proposition 3. Let f be a maximal local inverse signed edge domination function, for any $e = uv \in E$, if $f(e) = -1$, then there exist vertex $w \in N(u)$, such that $f(E(w)) \in \{0, 1\}$.

Proof. Let f be a maximal local inverse signed edge domination function, and assume that, there is a edge $e_0 = u_0v_0 \in E$ with $f(e_0) = -1$. For any vertex $w \in N(u_0)$, suppose $f(E(w)) \leq -1$. Define a new function $g : E \rightarrow \{1, -1\}$ and $g(e) = \begin{cases} 1; & e = e_0 \\ f(e); & e \neq e_0 \end{cases}$. When $w = v_0$, $g(E(w)) = f(E(w)) + 2 \leq 1$. For $w \neq v_0$,

$g(E(w)) = f(E(w)) \leq 1$. So g is also a local inverse signed edge domination function. Since $g > f$, the maximality of f is contradicted.

Note. If f is a signed edge domination function, this property is a sufficient and necessary condition.

Theorem 1. For every graph G , then $\tilde{\gamma}'_l(G) \leq \lceil n/2 \rceil$.

Proof Let f is a local inverse signed edge domination function of G , then for any $u \in V$, $f(E(u)) \leq 1$, hence $\sum_{u \in V} f(E(u)) \leq n$, note $2\tilde{\gamma}'_l(G) = \sum_{u \in V} f(E(u)) \leq n$, so $\tilde{\gamma}'_l(G) \leq n/2$ is valid. Because $\tilde{\gamma}'_l(G)$ is a positive integer, $\tilde{\gamma}'_l(G) \leq \lceil n/2 \rceil$.

Theorem 2. For any simple connected graph G with its order $m \geq 4$, then

$$\tilde{\gamma}'_l(G) \leq \lceil m + 4 - 2\sqrt{m+4} \rceil.$$

Proof. Let f be a $\tilde{\gamma}'_l(G)$ -function, $S_1 = \{e \in E \mid f(e) = 1\}$, $S_2 = \{e \in E \mid f(e) = -1\}$, $|S_1| = t$, $|S_2| = m-t$, hence $\tilde{\gamma}'_l(G) = 2t-m$.

For convenience, define a new graph G^* with $S_1 \cup S_2$ be its vertices set, two vertices in G^* are adjacent to each other if and only if they are adjacent to each other in G as edges. $S_1^* = \{v^* \in G^* \mid f(v^*) = 1\}$, $S_2^* = \{v^* \in G^* \mid f(v^*) = -1\}$, Let $E_{G^*}(S_1^*, S_2^*) = \{e = xy \mid x \in S_1^*, y \in S_2^*\}$. Because $f(E_G(v)) \leq 1$ for every $v \in V$ is true, hence $|N_G(x) \cap S_2^*| \geq 1$ for every $x \in S_1^*$, $|E_{G^*}(S_1^*, S_2^*)| \geq |S_1^*| = t$. There exist at least one vertex $y \in S_2^*$, such that y is at least adjacent to $\lceil t/(m-t) \rceil$ vertices in S_1^* . Hence, $|N_{G^*}[y] \cap S_2^*| \geq |N_{G^*}[y] \cap S_1^*| - 3 \geq \lceil t/(m-t) \rceil - 3$, $m-t = |S_2^*| \geq |N_{G^*}[y] \cap S_2^*| \geq \lceil t/(m-t) \rceil - 3$, Because the above inequality, we deduce that $t \leq m + 2 - \sqrt{m+4}$. Then $\tilde{\gamma}'_l(G) = 2t-m \leq m + 4 - 2\sqrt{m+4}$ is true. Because $\tilde{\gamma}'_l(G)$ is a positive integer,

$$\tilde{\gamma}'_l(G) \leq \lceil m + 4 - 2\sqrt{m+4} \rceil.$$

Theorem 3. For any connected graph G , Δ and δ be is largest and smallest degree, respectively,

$$\tilde{\gamma}'_l \leq \lceil (2\Delta - \delta + 3)m / (2\Delta + \delta + 1) \rceil.$$

Proof. Symbols are similar to theorem 2. For any $v \in V$, $f(E(v)) \leq 1$, hence $|E(v) \cap S_1| - |E(v) \cap S_2| \leq 1$, we know $|E(v) \cap S_1| + |E(v) \cap S_2| = d_G(v)$. $|E(v) \cap S_2| \geq (d_G(v)-1)/2 \geq (\delta-1)/2$.

From theorem 2, we know $E_{G^*}(S_1^*, S_2^*) = \sum_{x \in S_1^*} |N_{G^*}[x] \cap S_2^*| \geq \sum_{x \in S_1^*} (\delta - 1)/2 = t(\delta - 1)/2$. Note that $|S_2^*| = m - t$, so there exist at least one vertex y in S_2^* is adjacent to at least $\lceil t(\delta - 1)/(2m - 2t) \rceil$ vertices of S_1^* . Because $f(E(v)) \leq 1$, so y is adjacent to at least $\lceil t(\delta - 1)/(2m - 2t) \rceil - 3$ vertices of S_2^* . $\Delta(G^*) \geq d_{G^*}(y) \geq 2\lceil t(\delta - 1)/(2m - 2t) \rceil - 3$.
 $\Delta(G) \geq \frac{1 + \Delta(G^*)}{2} \geq \frac{1}{2}(1 + 2\left\lceil \frac{t(\delta - 1)}{2(m - t)} \right\rceil - 3) = \left\lceil \frac{t(\delta - 1)}{2(m - t)} \right\rceil - 1$. We derive that $t \leq (2m\Delta + 2m)/(\delta + 2\Delta + 1)$.
 $\tilde{\gamma}'_l(G) = 2t - m \leq \lceil (2\Delta - \delta + 3)m/(2\Delta + \delta + 1) \rceil$.

Theorem 4. G is a simple connected graph, m is the number of odd degree vertex, then $\tilde{\gamma}'_l(G) \leq \frac{m}{2}$.

Proof. Let f be a $\tilde{\gamma}'_l(G)$ -function, $\tilde{\gamma}'_l(G) = f(E)$. $S_1 = \{e \in E \mid f(e) = 1\}$, $E_2 = \{e \in E \mid f(e) = -1\}$, $\tilde{\gamma}'_l(G) = |S_1| - |S_2|$. Define two spanning subgraph G_1 and G_2 of G , $V(G_1) = V(G_2) = V(G)$, $E(G_1) = S_1$, $E(G_2) = S_2$. For any vertex $u \in V(G)$, $d^*(u) = d_{G_1}(u) - d_{G_2}(u)$, $\tilde{\gamma}'_l(G) = |E(G_1)| - |E(G_2)| = \frac{1}{2} \sum_{u \in V(G)} d^*(u)$.
 $2\tilde{\gamma}'_l(G) = \sum_{u \in V(G)} d^*(u)$. From the definition of local inverse signed edge domination function, $f(E(u)) \leq 1$. f is a $\tilde{\gamma}'_l(G)$ -function, know, when u is a even degree vertex, $f(E(u)) \leq 0$, when u is a odd degree vertex, $f(E(u)) \leq 1$. Let X be a set with even degree vertices, Y be a set with odd degree vertices, $V(G) = X \cup Y$, $X \cap Y = \emptyset$.

$$\tilde{\gamma}'_l(G) = \frac{1}{2} \sum_{u \in V(G)} d^*(u) = \frac{1}{2} \left(\sum_{w \in X} d^*(w) + \sum_{v \in Y} d^*(v) \right) = \frac{1}{2} \sum_{v \in Y} d^*(v) \leq \frac{1}{2} m$$

Theorem 4 is true.

Acknowledgments. This work is supported by the Educational Commission of Hebei Province (N0.ZH2011122, Z2011157), the Natural Science Foundation of Hebei province (N0.A2012408002) and the Youth Foundation of Langfang Theachers College (N0.LSZQ201106).

References

1. Favaron, O.: Signed domination in regular graphs. Discrete Math. 158, 287–293 (1995)
2. Dunbar, J., Hedetniemi, S., Henning, M.A.: Minus Domination in graphs. Discrete Mathematics 199, 35–47 (1999)

3. Xu, B.G.: On signed edge domination numbers of graphs. *Discrete Math.*, 179–189 (2011)
4. Xu, B.G.: On signed edge total domination of graphs. *Jounal of Mathematical Research and Exposition*, 8–12 (2007)
5. Xu, B.G.: On signed edge domination number of wheel graphs. *Jounal of Huadong Jiao Tong University*, 79–81 (2003)
6. Xu, B.G.: On signed edge total domination. *Jounal of Huadong Jiao Tong University*, 129–131 (2006)
7. Xu, B.G., Zeng, Y.: On the super bounds of signed edge domination numbers of graphs. *Jounal of Huadong Jiao Tong University*, 55–57 (2002)
8. Xu, B.G.: On the lower bounds of signed edge total domination numbers of graphs. *Jounal of Huadong Jiao Tong University*, 110–113 (2004)
9. Xu, B.G.: On signed edge domination numbers of graphs. *Jounal of Huadong Jiao Tong University*, 102–105 (2003)
10. Xu, B.G.: On reverse signed edge total domination numbers of graphs. *Jounal of Huadong Jiao Tong University*, 110–115 (2007)
11. Wang, H.C., Tong, J., Volkmann, L., Kiernan, K.P., Merey, S.A., Schroeder, M.W.: Signed Domination of Graphs and (0,1)-Matrices. *Combinatorics and Graphs* 531, 19–42 (2010)
12. Atapour, M., Tabriz, S.M.S., Khodkar, A.: Global signed total domination in graphs. *Publicaciones Mathematicae-Debrecen* 79, 7–22 (2011)

Forecast of Employment Based on Independent Component Analysis

Lu Tan and Han Zhang

Shandong University of Finance and Economics, School of Statistics and Mathematics,
Jinan, China

Abstract. To build a scientific model for predicting employment, indices which influence employment were selected from the five aspects, i.e. social and economic development, quality of workforce, development level of urban and rural area, living standard of workforce and working environment. By means of the independent component analysis (ICA), three main independent decomposition factors that reflect employment conditions, i.e. environmental factor, development factor and quality of workforce factor, were extracted from the indices. Based on the historical inertia and trend of the employment, the linear forecast model for the employment was built. By collecting the data of Shandong Statistical Yearbook from the year 1996 to 2008, a short-term forecast model for Shandong Province's employment was constructed. The result shows that the quality of workforce, industrial structure and income are essential factors to influence the employment of Shandong Province. It also provides useful information for the improvement of Shandong's employment situation.

Keywords: ICA, Independent Decomposition Factor, Linear Forecast.

1 Introduction

Reducing unemployment rate is one of the important targets of the macroeconomic regulation and control. Employment (or unemployment) issue is a major problem which is deeply concerned in the social and national economy. For a populous nation like China, employment problem becomes a more crucial livelihood issue that cannot be ignored. How to build a scientific and reasonable model of the employment in order to give a precise forecast for the relationship between current status of economic and social development and future employment situation becomes a hot issue in the employment research field. A suitable and accurate forecast model will offer useful suggestions for the government to establish rational politics about employment.

Nowadays, in order to solve the employment forecast problem, the main methods to construct a forecast model on employment are based on the time-series characteristics of the employment and the relevance between a specific economic development indicator and employment, including direct method, regression analysis and time series analysis. Direct method means calculating the employment elasticity directly, and then making predictions about the employment [1],[2], regression analysis focuses on building the linear (or could be transferred to linear) regression

models for forecasting [3],[4], while time series analysis is to produce the time series models [5]. Each method mentioned above has its advantages, whereas they all lack the ability to fully explore the direct influence that economy and social development have upon employment. Having the ability to obtain the independent components that can not only summarize the development of economy and society, but also influence employment, and to clarify the complicated relationships among numerous factors influencing employment become the key to success in forecasting employment. On the other hand, ICA, as one of the most effective methods in blind signal separation, can decompose the initial data into linear combinations of independent components, which sufficiently explains the deep reasons behind the phenomena. ICA has been widely applied in the fields of economy, finance, etc. with good results.

Starting off from the development level of economy and society and by means of ICA, the five aspects, social and economic development, quality of workforce, development level of urban and rural area, living standard of workforce and working environment, are decomposed independently. By ordering the decomposition components according to the relationship between these components and the employment, three main “independent decomposition factors”, environmental factor, development factor and quality of workforce factor, are achieved. After that, based on the historical inertia and the trend of the employment, a linear forecast is built for the predictions about the quantity of total employment as well as the quantities of employment by industries of Shandong Province, so as to provide some useful information for the improvement of Shandong’s employment situation.

2 Forecast Model for Employment by ICA

2.1 Construction of Index System of the Influencing Factors on Employment

Based on the previous research achievements on employment by scholars from China and abroad [6],[7],[8] and by using the “China Statistical Yearbook” and “Shandong Province Statistical Yearbook”, this paper selects twenty-four main indices that reflect the development status of economy and society from the five aspects (social and economic development, quality of workforce, development level of urban and rural area, living standard of workforce and working environment). These indices are chosen under the methods of merging similar data as well as focusing on important aspects, and are given full consideration of the completeness of data.

2.2 Construction of Independent Decomposition Factors

2.2.1 Principles of ICA

The basic model for ICA is the instantaneous linear mixture model, which considers the observed variables being linearly merged by many independent components [8]. Generally speaking, if the D observable random variables, denoted by X_1, X_2, \dots, X_D , can be expressed as the linear combinations by d independent components denoted by S_1, S_2, \dots, S_d , which is written as

$$X_i = \sum_{j=1}^d \alpha_{ij} S_j, \quad i = 1, 2, \dots, D, \quad (1)$$

where α_{ij} is the combination weight, $i = 1, 2, \dots, D$, $j = 1, 2, \dots, d$, or which can be written as a matrix $X = AS$, where A denotes a $D \times d$ weight matrix, $X = (X_1, X_2, \dots, X_D)^T$, $S = (S_1, S_2, \dots, S_d)^T$, then this model is an independent component model.

Table 1. Index System of the Influencing Factors on Employment

Serial Number	Influencing Factor	Index
	Conditions	Gross Domestic Product (GDP)
	For Social	Total Output of Primary Industry
	Economic Development	Total Output of Secondary Industry Total Output of Tertiary Industry Ratio of GDP by Primary Industry Ratio of GDP by Secondary Industry Ratio of GDP by tertiary Industry Total Export
	Quality of Workforce	Total Population Birth Rate Natural Growth Rate Graduates of Higher Education Institutes Expenditure on culture, education and science
	Development Level of Urban and Rural Areas	Disposable Income of Urban Households Net Income of Rural Households Urban Consumer Price Index Rural Consumer Price Index
	Living Standard of Workforce	Consumer Price Index of Shandong Province Hospital beds per 10,000 persons Number of Doctors per 10,000 persons Per Capita Construction Area of Building
	Working Environment	Volume of Soot Discharged Volume of Industrial Waste Water Discharged Volume of Sulphur Dioxide Discharged

In this model, the independent components are latent variables that cannot be observed directly, while at the same time the weight matrix A is unknown, so the basic target for ICA is to estimate S by analyzing the linear transformation of X ($S = WX$, where W is a $d \times D$ matrix), then to obtain A by solving the pseudo-inverse of W . Compared with PCA (Principal Component Analysis) and factor analysis, the feature for ICA is that it can achieve the maximization of non-Gaussianity of the decomposition components, meaning that the independent components depart more

from Gaussian distribution than any other observed variables. This paper makes analysis by using FastICA¹ algorithm, described in [9],[10].

2.2.2 Independent Decomposition of Influencing Factor Indices of Employment

Let $X = (X_1, X_2, \dots, X_{24})^T \in R^{24 \times 13}$ ($\text{rank}(X) = 12$) be the total observations of the indices of the influencing factors on employment, then the result after the independent decomposition is named $S = (S_1, S_2, \dots, S_{12})^T \in R^{12 \times 13}$, thus $S = WX$ (W is a 12×24 coefficient matrix).

2.2.3 Selection of the Independent Decomposition Factors

The independent decomposition factors S_i ($i = 1, 2, \dots, 12$) are ordered according to the correlation size ρ_i between S_i and the quantity of total employment C_0 as well as the quantities of employment by industries C_j ($j = 1, 2, 3$, which denotes the primary, secondary and tertiary industry, respectively), i.e.

$$\rho_i = \sum_{j=0}^3 | \text{corrcoef}(S_i, C_j) | , \quad (2)$$

where corrcoef denotes the correlation coefficient between two random variables. After that, the independent decomposition factors with a strong correlation are selected as the influential factors on employment, hence forecasting the employment.

2.3 Construction of Linear Forecast Model for Employment

In order to explain the result conveniently, we chose the independent decomposition factors² with a strong correlation as the direct social factors on the employment, meanwhile fully considered the historical inertia and trend of the employment issue, hence constructing the linear forecast model for the employment, t^3 to year $t+k$, i.e.

$$C_i(t+k) = \beta_{i0} + \sum_{j=1}^L \beta_{ij} \cdot S_j(t) + \sum_{j=0}^{T-1} \gamma_{ij} \cdot C_i(t-j) + \lambda_i \cdot (t+k) + \varepsilon_i(t+k) \quad (3)$$

where $i = 0, 1, 2, 3$; k denotes the prediction span; β_{ij} ($j = 1, 2, 3, \dots, L$) reflect the explanatory ability of the independent decomposition factors to the employment; L is the number of the chosen independent decomposition factors; γ_{ij} ($j = 0, 1, 2, \dots, T-1$) describe the historical inertia, i.e. autocorrelation, of the

¹ Algorithm software package is referred to <http://www.cis.hut.fi/projects/ica/fastica>.

² With the increasing number of independent decomposition factors being chosen, the variance of the error term will gradually decrease. In order for an easier explanation and analysis, we decided $L = 3$.

³ For convenience, let the initial year be $t = 1$.

employment; T represents the time span, which is simplified as $T = 1$; λ_i reflect the time-varying trend in the employment; $\varepsilon_i(t+k)$ denote prediction deviation, indicating the fluctuation which cannot be explained by the linear combination of independent decomposition factors, self-fluctuation, and time trend.

3 Forecast Model for Employment in Shandong Province, China

3.1 Calculation and Selection of Independent Decomposition Factors

We selected the data from the year 1996 to 2008 from the “Shandong Statistical Yearbook”. In order to have better use of the explanatory ability of the model parameters and to compare the results more easily, every X_i ($i = 1, 2, 3, \dots, 24$) was

standardized in the first place, written as $\frac{X_i - E(X_i)}{\sqrt{Var(X_i)}}$. After that, the fastICA

algorithm was applied to whitening and independently decomposing the standardized observing data. According to the results, we selected three independent components (still denoted by S_i , $i = 1, 2, 3$ for convenience) that have the strongest correlations with the employment as the independent decomposition factors. With the increasing number of independent decomposition factors being chosen, the variance of the error term will gradually decrease.

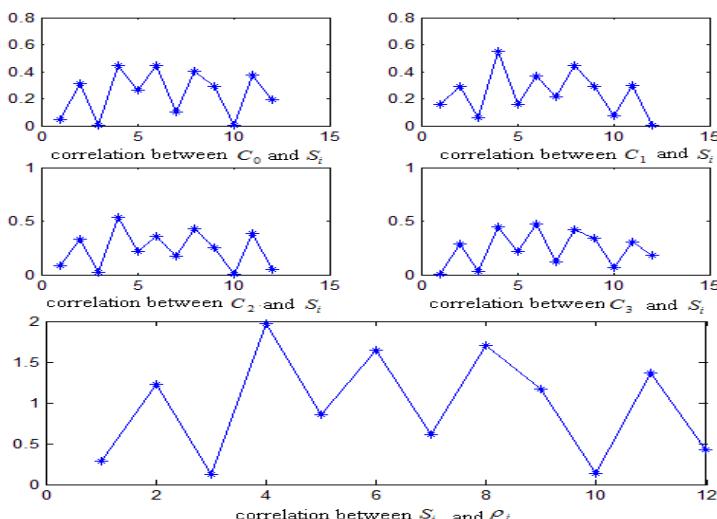


Fig. 1. Correlation between Employment and Independent Decomposition Factor

(Note: the horizontal axes denote the i^{th} independent decomposition factor, while the vertical axes denote the correlation size ρ_i).

Table 2. Coefficient matrix of three most correlated independent decomposition factors

Serial Number	S_1		S_2		S_3	
	Composition Index	Correlation Coefficient	Composition Index	Correlation Coefficient	Composition Index	Correlation Coefficient
1	0.11727	0.4033	-1.4981	0.4284	-0.7831	0.0922
2	0.17776	0.3949	-0.94256	0.3995	3.6929	0.1223
3	0.13104	0.4024	-1.7221	0.4261	-1.2312	0.0822
4	0.07829	0.405	-1.1976	0.4375	-0.9168	0.1037
5	1.2213	0.4946	0.49571	0.4006	2.5763	0.1398
6	-1.157	0.4398	0.99147	0.3106	4.3841	0.1793
7	-1.1193	0.3245	-1.0068	0.0538	1.4275	0.1223
8	0.32834	0.3949	-1.3006	0.4144	1.2237	0.1194
9	-1.0606	0.4823	1.3441	0.397	4.3779	0.0672
10	0.033362	0.2124	0.41445	0.1814	-1.3197	0.462
11	0.079833	0.1716	-1.5762	0.0172	0.21541	0.5662
12	1.7904	0.3805	1.4525	0.4348	-5.5762	0.071
13	0.58535	0.4027	-3.8435	0.4425	-2.0375	0.0563
14	0.20992	0.4293	-0.8020	0.4187	0.66884	0.0692
15	-0.0597	0.4281	0.13629	0.4344	-0.0795	0.0754
16	0.075344	0.4484	-1.777	0.4443	-1.3824	0.0459
17	-0.02512	0.3556	0.84686	0.3661	0.63912	0.1186
18	-0.81447	0.3857	0.91916	0.3635	3.9293	0.1308
19	-2.6923	0.3919	5.533	0.439	-0.3037	0.2194
20	0.91231	0.1335	-0.7674	0.485	-0.7836	0.2579
21	0.03984	0.4172	1.4896	0.4046	-1.3774	0.2065
22	-0.79252	0.6139	-2.9033	0.3788	-1.229	0.1365
23	0.62566	0.2998	-0.1135	0.2995	-1.4738	0.1559
24	-0.81022	0.5667	0.50745	0.3242	1.3509	0.0401

Note: the indices X_i in table 2 have been standardized, i.e., $EX_i = 0$, $Var(X_i) = 1$, $i = 1, 2, \dots, 24$.

From the analysis of the composition of the independent decomposition factors S_i , we can conclude that S_1 reflects the aspect of the influencing factors relating with the working and living environment, named as the Environmental Factor; S_2 reflects the aspect of social and economic development, named as the Development Factor; S_3 reflects the aspect of the quality of workforce, thus named as the Quality of Workforce Factor.

3.2 Construction and Analysis of Forecast Model of Employment

By using the regression function in Matlab, we put the Environmental Factor $S_1(t)$, the Development Factor $S_2(t)$, the Quality of Workforce Factor $S_3(t)$, as well as the Employed Population $C_i(t)$ and $C_i(t+k)$ into the linear forecast model of employment, resulting in table 3 and 4 (Due to space constraints, we just displayed the results at $k=1$ and $k=5$).

Table 3. Model test results and parameter values at $k = 1$

Employment(in 10,000,000)	Total	Primary Industry	Secondary Industry	Tertiary Industry
Model Parameter	β_0	0.9166	0.751	0.1732
	β_1	-0.0045	-0.0285	0.0249
	β_2	0.0129	0.0385	-0.0146
	β_3	-0.0174	-0.0402	0.0131
	β_4	0.829	0.7824	0.7911
	β_5	0.0179	-0.0316	0.029
Residual Variance		0.0006	0.0038	0.0047
R^2		0.9968	0.9651	0.9695
P Value		<1e-6	0.0003	0.0002
				<1e-6

Table 4. Model Test Results and Parameter Values at $k = 5$

Employment(in 10,000,000)	Total	Primary Industry	Secondary Industry	Tertiary Industry
Model Parameter	β_0	-0.2061	5.4514	1.3772
	β_1	-0.00145	-0.0076	0.0446
	β_2	-0.0027	-0.0085	-0.0264
	β_3	-0.0185	0.0035	0.0159
	β_4	1.0779	-0.8917	-0.2014
	β_5	0.0444	-0.0991	0.1177
Residual Variance		0.0003	0.0007	0.0020
R^2		0.9986	0.9960	0.9919
P Value		0.0036	0.0101	0.0202
				0.0103

We can conclude from the Table 3 and 4 that the forecast residual is small whether it's at $k = 1$ or $k = 5$, or whether it's for the total employment or that of each industry. Besides, R^2 are approximately at 0.99, which means a sufficient linear explanation for the employment. P values are also quite small. All the results above demonstrate that the linear forecast model based on the independent components can fully explain the driving mechanism of employment and the reasons for fluctuation. Moreover, it is clear from the parameter values that because of the different industries and prediction spans, the independent decomposition factors have different explanatory abilities for the employment drive. For instance, the explanatory ability of $k = 5$ is much inferior to that of $k = 1$, meaning that as time increases, the present influencing factors on employment would be gradually weaker. Another example is the employment model of the tertiary industry at $k = 5$, where the industry factor is a weak driving force for employment, while the change in time can explain well for the employment fluctuations.

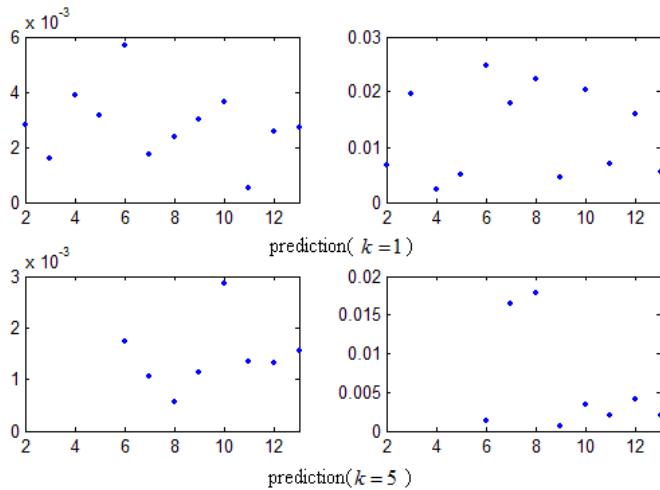


Fig. 2. Figures of Relative Error of Prediction Model on Employment

(Note: the horizontal axes denote the year, while the vertical axes denote the predicted relative error. The two graphs on the left side are for the total employment, and those on the right are for the employment of tertiary industry.)

Besides, we can find that the R^2 at $k=1$ is slightly smaller than that at $k=5$, while P value is bigger. It demonstrates that the independent decomposition factors have a lagged effect on employment, meaning the policy adjustment on employment in year t would have a bigger influence on the employment of year $t+k$ ($k > 2$) than that of year $t+1$.

The results from the Table 3 and 4 demonstrate the rationality of the model that has been constructed, i.e. the driving force of the independent decomposition factors for employment. From Figure 2 we can find that most of the relative errors can be controlled within 2%, indicating a well explanation for the source of employment fluctuations.

4 Conclusions

ICA is a latent variable analysis which has gradually become popular since 1990s. Under certain conditions, it can effectively extract initial signals that are independent of each other from the mixed observation signals, thus obtaining the deep reasons behind the phenomena. Employment variation results from the combined effect of social and economic development, quality of workforce, development level of urban and rural area, living standard of workforce and working environment, which means that employment can be treated as a mixed signal. One important feature for ICA is that it can extract initial signals that are independent of each other from several mixed signal series, so as to find deeper reasons that affect employment.

By using ICA, this paper extracted the influencing factors of employment, and found that the quality of workforce, industrial structure and income are core factors. It also deeply analyzed the driving mechanism of employment and the potential essence of fluctuation, so as to provide useful suggestions for the improvement of employment situation in Shandong Province, China.

Acknowledgments. This work is supported by National Social Science Foundation (No.11BGL072) and Natural Science Foundation of Shandong Province (No. ZR2009HM007 and No.ZR2011AL022).

References

1. Cai, F., Du, Y., et al.: Employment Elasticity, NAIRU and Macroeconomic Policies- Why economic growth did not bring the dominant employment. *Economic Research Journal* (2004)
2. Ding, S.: How Much is Chinese Employment Elasticity? - And the Financial Crisis Impact on Employment. *Management World* (2009)
3. Li, T., Liu, J., et al.: Mathematical Model and Prospect Forecast of Employment in China. *Mathematics* (2010)
4. Xiao, Y., Zhou, Q., et al.: Quantitative Analysis on China's Long-term and Short-term Factors Affecting the Employment. *Statistics and Decision* (2010)
5. Xiao, D., Fu, K., et al.: The Employment Growth Forecast Model Based on Box-Jenkins Modeling Method. *Statistics and Decision* (2008)
6. Zhang, D.: Early Warning Research of Labor and Employment and Unemployment in the Process of China's Rapid Economic Growth. PhD thesis, Fudan University (2007)
7. Peng, X., Zhong, X.: Causes of Coexistence on China's Current High Economic Growth and Rising Unemployment. *Macroeconomic Management* (2009)
8. Hyvärinen, A., Karhunen, et al.: *Independent Component Analysis*. John Wiley & Sons Inc., New York (2001)
9. Hyvärinen, A., Oja, E.: *Independent Component Analysis: Algorithms and Applications*. Neural Networks (2000)
10. Taiji, S.: Least-squares Independent Component Analysis. *Neural Computation* (2011)

Complete Convergence for NA Random Sequence

Yourong Wang^{1,*}, Yanli Liu¹, and Yili Tan²

¹ Department of Basic, Tangshan College, Hebei Tangshan 063000, China

² College of Science, Hebei United University, Hebei Tangshan 063009, China
{yourong1214, tanyili_2002}@163.com, ly17937@126.com

Abstract. The classical complete convergence theorem concerns the arithmetic means which is a regular method of summability. In this paper, to obtain the main results, a new large class of summability methods is introduced. The results for complete convergence for negatively associated random variable sequence are obtained. To investigate this results, by restricting the moment conditions and use a new method of summability. Then the result of the complete convergence for NA random variables sequences are obtained by applying the Kolmogorov-type inequality for NA random variable sequence.

Keywords: Negatively Associated, Complete Convergence, Kolmogorov-Type Inequality.

1 Introduction

Since the concept of complete convergence was introduced by Hsu and Robbins (1947), there are many people who devote the study to complete convergence for i.i.d. random variables.

In this paper, by restricting the moment conditions, we obtain some results for NA random variable.

Definition 1. Let $\{X_n, n \geq 1\}$ be a sequence of random variables. Let $S_n = \sum_{i=1}^n X_i$ be the partial sum of $\{X_n, n \geq 1\}$.

A finite family $\{X_1, \dots, X_n\}$ is said to be negatively associated (abbreviated to NA) if for any disjoint subsets $A, B \subset \{1, 2, \dots, n\}$ and any real coordinate-wise non-decreasing f, g , we have

$$\text{Cov}(f(X_i, i \in A), g(X_j, j \in B)) \leq 0.$$

Infinite family of random variables is NA if every finite subfamily is NA. This concept was introduced by Joag-Dev and Proschan (1983).

To prove the theorem we need the following Kolmogorov-type inequality for NA random variable sequence which was introduced by Tasos C.Christofides & Eutichia Vaggelatou (2004).

* Corresponding author.

Lemma let $\{X_n, n \geq 1\}$ be a NA random variable sequence, let $EX_n = 0$, $EX_n^2 < \infty$ then for $\forall \varepsilon > 0$,

$$P(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon) \leq 2\varepsilon^{-2} \sum_{i=1}^n EX_i^2$$

$$P(\max_{1 \leq k \leq n} S_k \geq \varepsilon) \leq \varepsilon^{-2} \sum_{i=1}^n EX_i^2$$

In this paper, a large class of summability methods is considered.

Definition 2. Let g be a positive increasing function and h a positive function such that $\varphi(y) \equiv g(y)h(y)$ satisfies the following conditions:

- (i) For some $d \geq 0$, φ is strictly increasing on $[d, +\infty)$ with range $[0, +\infty)$,
- (ii) There exist C and a positive integer k_0 such that $\varphi(y+1)/\varphi(y) \leq C$, $y \geq k_0$,
- (iii) There exist constants a and b such that

$$\varphi(s)^2 \int_s^{+\infty} \frac{1}{\varphi(x)^2} dx \leq as + b, \quad s > d.$$

For h and g as above, the (h,g) -transform of sequence $\{\xi_n\}$ is given by

$$\sigma_n(x) = \frac{1}{g(n)} \sum_{k=1}^n \frac{1}{h(k)} \xi_k, \quad n = 1, 2, \dots$$

If $\sigma_n(x) \rightarrow \xi (n \rightarrow \infty)$, then we say that $\{\xi_n\}$ summable (limitable) to ξ by the method (h,g) , and write $(h,g)-\lim \xi_k = \xi$.

Definition 2 is presented by Ryszard Jajte [5]. It should be stressed here that the above class of sequence transformations includes several summability methods such as arithmetic means [$h(y) = 1$, $g(y) = y$] or logarithmic means [$h(y) = y$, $g(y) = \log y$] which also embraces nonregular transformations.

2 Main Result

2.1 Theorem

The results for Complete convergence for negatively associated random variable sequence is obtained, by restricting the moment conditions and use the new methods of summability.

Theorem let $\{X, X_n, n \geq 1\}$ be a NA random variable sequence with identical distribution, let $h(n)$ be a positive decreasing function, $g(n)$ be a positive increasing function and $\varphi(n) \equiv h(n)g(n)$ be as in the definition 1, let $S_k = \sum_{i=1}^k \frac{X_i}{h(i)}$,

Such that

$$E[\varphi^{-1}(|X|)] < \infty, \text{then } \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon g(n)\right) < \infty.$$

Proof

Let $X'_{kn} = X_k I(|X_k| \leq \varphi(n)) + \varphi(n) I(|X_k| > \varphi(n))$, and $T_n = \sum_{k=1}^n \frac{X'_{kn}}{h(k)}$,

$$\text{Since } \left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon g(n) \right) \subset \left(\max_{1 \leq k \leq n} |T_k| \geq \varepsilon g(n) \right) \cup \left(\bigcup_{k=1}^n X'_{kn} \neq X_{kn} \right)$$

$$\begin{aligned} \text{So that } & \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon g(n)\right) \\ & \leq \sum_{n=1}^{\infty} \frac{1}{n} \left[P\left(\max_{1 \leq k \leq n} |T_k| \geq \varepsilon g(n)\right) + P\left(\bigcup_{k=1}^n X'_{kn} \neq X_{kn}\right) \right] = I_1 + I_2 \end{aligned}$$

Since

$$\begin{aligned} P\left(\bigcup_{k=1}^n X'_{kn} \neq X_{kn}\right) & \leq \sum_{k=1}^n P(X'_{kn} \neq X_{kn}) \\ & = \sum_{k=1}^n P(|X_k| \geq \varphi(n)) \\ & = \sum_{k=1}^n P(|X| \geq \varphi(n)) \\ & = n P(|X| \geq \varphi(n)) \end{aligned}$$

So

$$\begin{aligned} I_2 & \leq \sum_{n=1}^{\infty} \frac{1}{n} n P(|X| \geq \varphi(n)) \\ & = \sum_{n=1}^{\infty} P(|X| \geq \varphi(n)) \\ & = \sum_{n=1}^{\infty} P(\varphi^{-1}(|X|) \geq n) \\ & \leq E[\varphi^{-1}(|X|)] \\ & < \infty \end{aligned}$$

And since

$$\begin{aligned} I_1 & = \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} |T_k| \geq \varepsilon g(n)\right) \\ & = \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \frac{X'_{in}}{h(i)} \right| \geq \varepsilon g(n)\right) \\ & \leq \sum_{n=1}^{\infty} \frac{1}{n} \left[P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \frac{X_i I(|X_i| \leq \varphi(n))}{h(i)} \right| \geq \frac{\varepsilon}{2} g(n)\right) \right] \end{aligned}$$

$$+P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \varphi(n) \frac{I(|X_i| > \varphi(n))}{h(i)} \right| \geq \frac{\varepsilon}{2} g(n)\right)$$

$$= I'_1 + I''_1$$

Then from the lemma, we have

$$\begin{aligned} I''_1 &\leq \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k I(|X_i| > \varphi(n)) \right| \geq \frac{\varepsilon}{2}\right) \\ &\leq \sum_{n=1}^{\infty} \frac{1}{n} 2 \left(\frac{\varepsilon}{2}\right)^{-2} \sum_{i=1}^n E[I(|X_i| > \varphi(n))]^2 \\ &= 2 \left(\frac{\varepsilon}{2}\right)^{-2} \sum_{n=1}^{\infty} \frac{1}{n} \sum_{i=1}^n P(|X_i| > \varphi(n)) \\ &= 2 \left(\frac{\varepsilon}{2}\right)^{-2} \sum_{n=1}^{\infty} P(|X| > \varphi(n)) \\ &\leq 2 \left(\frac{\varepsilon}{2}\right)^{-2} E[\varphi^{-1}(|X|)] \\ &< \infty \end{aligned}$$

From the lemma and since $h(n)$ is a positive decreasing function,

So that

$$\begin{aligned} I'_1 &\leq \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} \left| \sum_{i=1}^k \frac{X_i I(|X_i| \leq \varphi(n))}{h(i)} \right| \geq \frac{\varepsilon}{2} g(n)\right) \\ &\leq \sum_{n=1}^{\infty} \frac{1}{n} 2 \left(\frac{\varepsilon}{2} g(n)\right)^{-2} \sum_{i=1}^n E\left[\frac{X_i I(|X_i| \leq \varphi(n))}{h^2(i)}\right]^2 \\ &\leq 2 \left(\frac{\varepsilon}{2}\right)^{-2} \sum_{n=1}^{\infty} \varphi^{-2}(n) E X^2 I(|X| \leq \varphi(n)) \\ &\leq 2 \left(\frac{\varepsilon}{2}\right)^{-2} E \sum_{n=1}^{\infty} \varphi^{-2}(n) X^2 I(|X| \leq \varphi(n)) \\ &\leq 2 \left(\frac{\varepsilon}{2}\right)^{-2} E[\varphi^{-1}(|X|)] \end{aligned}$$

Since $E[\varphi^{-1}(|X|)] < \infty$, then $I'_1 < \infty$.

2.2 Corollary

Corollary. let $\{X, X_n, n \geq 1\}$ be a NA random variable sequence with identical distribution, let $EX_n = 0$, and let $h(n) = 1, g(n) = n^{1/a}$ ($0 < a < 2, a \neq 1$) be a positive increasing function and $\varphi(n) \equiv h(n)g(n)$ be as in the definition 1, such that

$$E[\varphi^{-1}(|X|)] < \infty,$$

then

$$\sum_{n=1}^{\infty} \frac{1}{n} P\left(\left|\frac{S_n}{n^{1/a}}\right| \geq \varepsilon\right) < \infty.$$

Proof

From the theorem

$$\sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon n^{1/a}\right) < \infty$$

So

$$\sum_{n=1}^{\infty} \frac{1}{n} P\left(\left|\frac{S_n}{n^{1/a}}\right| \geq \varepsilon\right) \leq \sum_{n=1}^{\infty} \frac{1}{n} P\left(\max_{1 \leq k \leq n} |S_k| \geq \varepsilon n^{1/a}\right) < \infty$$

References

1. Joag-Dev, K., Proschan, F.: Negative association of random variables with applications. *Ann. Statistic.* 11, 286–295 (1983)
2. Hsu, P.L., Robbins, H.: Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. USA* 33, 25–31 (1947)
3. Shao, Q.-M., Su, C.: The Law of iterated logarithm for negatively associated random variables. *Process. Appl.* 83, 139–148 (1999)
4. Lin, Z., Lu, C., Su, Z.: Foundations of probability limit theory. *College Mathematics Journal*, 98–106 (1999)
5. Christofides, T.C., Vaggelatou, E.: A connection between supermodular ordering and positive/negative association. *Journal of Multivariate Analysis* 88, 138–151 (2004)
6. Liu, L., Sheng, S.: On Sufficient and Necessary Condition of a Strong Law of Large Numbers for Negatively Associated Random Variables. *Chinese Journal of Applied Probability and Statistics* 24(4) (2008)
7. Zhou, S.-N., Ming, R.-X., Huang, L.: Complete Convergence for Randomly Weighted Sums of Negatively Associated Random Variables. *Journal of Jiangxi Normal University (natural sciences edition)*, 33(6) (2009)
8. Liang, H., Zhang, J.: Strong Convergence for Weighted Sums of Negatively Associated Arrays. *Chinese Annals of Mathematics, Series B* 31(2) (2010)
9. Xing, G., Yang, S.: Exponential Inequalities for Negatively Associated Random Variables. *Acta Mathematica Scientia*, 29(6) (2009)
10. Christofides, T.C., Hadjikyriakou, M.: Exponential inequalities for N-demimartingales and negatively associated random variables. *Statistics & Probability Letters* 79(19) (2009)

A Remark on Distance Regular Graph with $k = 10, a_1 = 1$

Junhong Ma and Bing Han

HeBei United University, School of Science,
Tangshan, Hebei, China
mjhgsg@163.com

Abstract. Distance-regular graph is a kind of highly regular graph. It is studied actively in recent years. The classification is one of the important contents in the study of Distance-regular graph. In this thesis we discuss the distance-regular graph with $k = 10, a_1 = 1$ by mean of intersection diagrams, circuit chasing techniques and properties of distance-regular graphs. We prove the following conclusion: Let Γ be a distance-regular graph of $k = 10, a_1 = 1$, if $c_{r+2} = 4$ and $a_{r+2} = 5$, then $c_d = 10$. It is useful to a classification of distance-regular graph of $k = 10, a_1 = 1$.

Keywords: Distance-regular graphs, intersection diagrams, intersection numbers, C_i -graph, vertex type.

1 Introduction

The definition of Distance-regular graph is put forward by British mathematician N.L.Biggs in the 1970s; it is extension of Distance-transitive graphs in algebra combination.

Distance-regular graph is highly regular; it has deep contact with finite group, combinatorial design, finite geometries, encoding, etc. Distance-regular graphs have been an important branch of algebra combination.

The classification is the core content of the study of Distance-regular graphs. The classification of Distance-regular graphs with valency 3 is completed by T.Ito, N.L.Biggs and A.G.Boshier in the 1980s; the classification of Distance-regular graphs with valency 4 is completed by A.E.Brouwer and J.Koolen with computer in 2002; Japanese mathematician A.Hiraki and H.Suzuki completed classification of Distance-regular graphs of $k = 6$ and $a_1 = 1$ by mean of circuit chasing techniques and characteristic values. And yet the classification of Distance-regular graphs of $k \geq 7$ or $a_1 > 1$ are not fully known.

In this article we discuss the parameters of distance-regular graph of $k = 10, a_1 = 1$. We hope that the conclusion of this article is useful to a classification of distance-regular graph of $k = 10, a_1 = 1$.

2 The Main Result in This Paper

Theorem: Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$, if $c_{r+2} = 4$ and $a_{r+2} = 5$, then $c_d = 10$.

3 Preliminaries

Definition 1. Let Γ be a connected graph with vertex set X and edge set E . Let u and v be vertices of Γ such that $\partial(u, v) = h$. Γ is a distance-regular graph if the number $|\Gamma_i(u) \cap \Gamma_j(v)|$ depend only on h, i, j and do not depend on the choices of individual vertices u and v of Γ .

More information about distance-regular graph will be found in A.E.Brouwer, A.M.Cohen, and A.Neumaier [6].

Let $\Gamma = (X, E)$ be a graph and $u \in \Gamma$, $\Gamma_i(u) = \{v \in \Gamma \mid \partial_\Gamma(u, v) = i\}$ and $\Gamma(u) = \Gamma_1(u)$. For vertices u, v of Γ with $\partial(u, v) = i$, let

$$c_i = |C_i(u, v)| = |\Gamma_{i-1}(u) \cap \Gamma(v)|$$

$$a_i = |A_i(u, v)| = |\Gamma_i(u) \cap \Gamma(v)|$$

$$b_i = |B_i(u, v)| = |\Gamma_{i+1}(u) \cap \Gamma(v)|$$

These numbers c_i , a_i and b_i are called intersection numbers. They play important roles in the study of distance-regular graphs.

More information about intersection numbers will be found in A.E.Brouwer, A.M.Cohen, and A.Neumaier [6].

Definition 2. Let u and v be adjacent vertices in Γ and $D_j^i = \Gamma_i(u) \cap \Gamma_j(v)$.

The intersection diagram with respect to (u, v) is the collection $\{D_j^i\}_{i,j}$ ($0 \leq i, j \leq d$) with lines between them. We draw a line if there is possibility of the existence of edges, and we erase the line when we know that there is no edge between D_j^i and D_t^s .

More information about intersection diagram will be found in A.E.Brouwer, A.M.Cohen, and A.Neumaier [6].

4 Proof of the Theorem

Lemma 1. Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. If $c_{r+1} = 2$, then any $x \in D_{r+1}^{r+1}$, $e(x, D_r^r) \leq 1$. In particular, if $a_{r+1} = 2$, then $D_{r+1}^{r+1} \subset \Gamma_r(\gamma)$.

Proof: Assume there exists $x \in D_{r+1}^{r+1}$ such that $e(x, D_r^r) > 1$. Set $y_r, z_r \in \Gamma(x) \cap D_r^r$. Since $c_{i+1} = 1$ for all $1 \leq i \leq r-1$, $\{y_{r-i}\} = \Gamma(y_{r-i+1}) \cap D_{r-i}^{r-i}$, $\{z_{r-i}\} = \Gamma(z_{r-i+1}) \cap D_{r-i}^{r-i}$ and $y_{r-i} \neq z_{r-i}$. In particular, $y_1 \neq z_1$ and $y_1, z_1 \in D_1^1$. This contradicts that $a_1 = 1$. Hence for any $x \in D_{r+1}^{r+1}$, we have $e(x, D_r^r) \leq 1$.

If $a_{r+1} = 2$, we count the number of edges between D_{r+1}^{r+1} and D_r^r in two ways. Since $e(x, D_{r+1}^{r+1}) = b_r$ for any $x \in D_r^r$, $e(D_r^r, D_{r+1}^{r+1}) = |D_r^r| \cdot b_r$. On the other hand, $e(D_r^r, D_{r+1}^{r+1}) = |D_{r+1}^{r+1}| \cdot m$. Hence $|D_r^r| \cdot b_r = |D_{r+1}^{r+1}| \cdot m$. Thus we have $D_{r+1}^{r+1} \subset \Gamma_r(\gamma)$.

Lemma 2. Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. If $c_{r+1} = 2$ and $a_{r+1} = 2$, then $e(D_{r+1}^{r+1}, D_{r+2}^{r+1}) = e(D_{r+1}^{r+1}, D_{r+1}^{r+2}) = 0$.

Proof: Assume there exist $x \in D_{r+1}^{r+1}, y \in D_{r+2}^{r+1}$ such that $x \sim y$. We have $\vec{x} = (1, 0, 1)$ and $\vec{y} = (1, 1, 2)$ from lemma 1. On the other hand, we note that there is not the vertex type of $(1, 1, 2)$ in Γ . Therefore $e(D_{r+1}^{r+1}, D_{r+2}^{r+1}) = 0$.

Similarly, we have $e(D_{r+1}^{r+1}, D_{r+1}^{r+2}) = 0$.

Lemma 3. Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. If $c_{r+2} = 4$, then $a_{r+2} \neq 6$.

Proof: Suppose that $a_{r+2} = 6$, then $D_{r+2}^{r+2} \neq \varphi$ and $d = r+2$. It is clear that there may exist only two vertex types of $(2, 1, 2)$ and $(2, 2, 2)$ in D_{r+2}^{r+2} . We have there exists the vertex type of $(2, 1, 2)$ in D_{r+2}^{r+2} from lemma 1 and lemma 2.

We assert that there exists the vertex type of $(2, 2, 2)$ in D_{r+2}^{r+2} at that time. In fact, suppose not, then there exists only vertex type of $(2, 1, 2)$ in D_{r+2}^{r+2} . Take any $x \in D_{r+2}^{r+2}$, we have $\vec{x} = (2, 1, 2)$. Lemma 2 implies that $C_{r+1}(\gamma, x) \subset D_{r+1}^{r+1}$. We have $e(x, D_{r+1}^{r+1}) = 2$ from lemma 1 and $c_{r+1} = 2$. Now we count the number of edges between D_{r+1}^{r+1} and D_{r+2}^{r+2} . We have $e(D_{r+1}^{r+1}, D_{r+2}^{r+2}) = 6 \cdot 8^r$. On the other hand, $e(D_{r+1}^{r+1}, D_{r+2}^{r+2}) = 9 \cdot 8^r$. This is a contradiction.

Take $\vec{y} = (2, 2, 2)$ in D_{r+2}^{r+2} , we have $e(y, D_{r+1}^{r+1}) = 0$ from lemma 1. Since $c_{r+2} = 4$, $e(y, D_{r+2}^{r+1}) = e(y, D_{r+1}^{r+2}) = 4$. Set $\{y_1, y_2, y_3, y_4\} = \Gamma(y) \cap D_{r+2}^{r+1}$, $\{z_1, z_2, z_3, z_4\} = \Gamma(y) \cap D_{r+1}^{r+2}$.

Consider $A_{r+2}(\gamma, y)$, we have $\{y_1, y_2, y_3, y_4, z_1, z_2, z_3, z_4\} \subset A_{r+2}(\gamma, y)$. This contradicts $a_{r+2} = 6$.

Lemma 4. Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. If $b_{d-1} = 1$ and $c_d = 8$, then $k_d \equiv 0 \pmod{3}$.

Proof: Let $\partial(\alpha, \beta) = d$, $D_j^i = \Gamma_i(\alpha) \cap \Gamma_j(\beta)$. We consider the intersection diagram with respect to (α, β) . Since $c_d = 8$ and $a_d = 2$, $|D_1^{d-1}| = 8, |D_1^d| = 2$. Since $b_{d-1} = 1$, $e(D_1^{d-1}, D_1^d) = 0$. We note that $\Gamma(\beta) \cong 5K_2$, hence $D_1^{d-1} \cong 4K_2$ and $D_1^d \cong K_2$. Set $\{\beta_1, \beta_2\} = D_1^d$ and $\beta_1 \sim \beta_2$. It is clear that $\{\beta, \beta_1, \beta_2\}$ is a clique in $\Gamma_d(\alpha)$.

Take any $\beta' \in \Gamma_d(\alpha)$ such that $\beta' \neq \beta, \beta_1, \beta_2$. We have $\{\beta'_1, \beta'_2\} = \Gamma(\beta') \cap \Gamma_d(\alpha)$ by $a_d = 2$. The intersection diagram with respect to (α, β') implies that $\{\beta', \beta'_1, \beta'_2\}$ is an other clique. Since $c_d = b_1 = 8$, $e(D_1^d, D_2^d) = 0$. So we have $\{\beta, \beta_1, \beta_2\}$ and $\{\beta', \beta'_1, \beta'_2\}$ are two distinct cliques. Thus we have $k_d \equiv 0 \pmod{3}$.

Lemma 5. Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. If $c_{r+2} = 4$, $a_{r+2} = 5$ and $d = r + 3$, then $c_d = c_{r+3} \neq 8$.

Proof: Suppose not. Since $d = r + 3$ and $b_{r+2} = b_{d-1} = 1$, $k_d \equiv 0 \pmod{3}$ from lemma 4. On the other hand, $k_d = \frac{10 \cdot 8^r \cdot 1}{1^r \cdot 4 \cdot 8}$ is not a multiple of 3. This is a contradiction.

Let Γ be a distance-regular graph of $b_{r+2} = b_{d-1} = 1$. Set $(c_{r+1}, a_{r+1}, b_{r+1}) = (2, 2, 6)$. If $c_{r+2} = 4$ and $a_{r+2} = 5$. C_{r+2} -graph is a coclique, so C_{r+3} -graph is a union of K_2 . Therefor we have $c_{r+3} = 8$ or $c_{r+3} = 10$.

If $d = r + 3$, then $c_d = 10$ is clear from lemma 5.

If $d \geq r + 4$, then $(c_{r+3}, a_{r+3}, b_{r+3}) = \dots = (c_{d-1}, a_{d-1}, b_{d-1}) = (8, 1, 1)$. We have $c_d = 10$ from lemma 4.

Now we complete the proof of theorem.

References

1. Hiraki, A., Nomura, K., Suzuki, H.: Distance-regular graphs of valency 6 and $a_1=1$. *J. Alge Combin.* 11, 101–134 (2000)
2. Yamazaki, N.: Distance-regular graphs with $\Gamma(x) \cong 3.K_{a+1}$. *Europ. J. Combin.* 16, 525–536 (1995)
3. Bannai, E., Ito, T.: On distance-regular graphs with fixed valency, II. *Graphs and Combin.* 4, 219–228 (1988)
4. Bannai, E., Ito, T.: On distance-regular graphs with fixed valency, III. *Journal of Algebra* 107, 43–52 (1987)
5. Bannai, E., Ito, T.: On distance-regular graphs with fixed valency, IV. *European J. Combin.* 10, 137–148 (1989)
6. Hiraki, A.: Circuit chasing technique for a distance-regular graph with c_{2r+1} . *Kyushu J. Math.* 49, 197–291 (2010)

7. Bannai, E., Ito, T.: On distance-regular graphs with fixed valency. *Graphs and Combin.* 3, 95–109 (1987)
8. Brouwer, A.E., Cohen, A.M., Neumaier, A.: *Distance-regular graphs*. Springer, Heidelberg (2010)
9. Hiraki, A., Suzuki, H.: On distance-regular graphs of $b_1 = c_{d-1}$. *Math. Japonica* 37, 751–756 (2012)
10. Suzuki, H.: *Local and antipodal structures of distance-regular graphs*. Springer, Heidelberg (2010)

Information Entropy, Similarity Measure and Inclusion Measure of Intuitionistic Fuzzy Sets

Qiansheng Zhang¹, Fuchun Liu², Lihua Wu³, and Shihua Luo⁴

¹ School of Informatics, Guangdong University of Foreign Studies,
Guangzhou 510420, China

² Faculty of Computer, Guangdong University of Technology, Guangzhou 510006, China

³ College of Informatics, South China Agricultural University, Guangzhou 510642, China

⁴ Faculty of Statistics, Jiangxi University of Finance and Economics,
Nanchang 330013, China

{Qiansheng Zhang, Zhqiansh01}@126.com

Abstract. This paper introduces a new axiomatic definition of the inclusion measure for intuitionistic fuzzy sets (IFSs, for short). The close relationships among information entropy, similarity measure, and inclusion measure of IFSs are then discussed in detail. Also, it investigates some important theorems by which the information entropy, similarity measure and inclusion measure of IFSs can be deduced easily based on their axiomatic definitions. Simultaneously, some new formulae to calculate information entropy, similarity measure and inclusion measure of IFSs are put forward.

Keywords: Intuitionistic fuzzy set, Information entropy, Similarity measure, Inclusion measure.

1 Introduction

As is well known, the information entropy and similarity measure, and inclusion measure are three important issues in fuzzy set theory, which have been widely applied to pattern recognition, cluster analysis, image processing and decision making. With respect to intuitionistic fuzzy set initialed by Atanassov [1], the three concepts still play significant role in many deployed systems and application problems involving intuitionistic fuzzy sets. The similarity measure of IFSs indicates the similar degree of two IFSs and plays an important role in pattern recognition, approximate reasoning and decision making. For instance, in [7] Szmudit presented a similarity measure of IFSs for supporting medical diagnostic reasoning. Later, Ye [14] proposed some cosine similarity measures for IFs. Hung [6] also presented a similarity measure of IFSs based on Hausdorff metric and applied it to pattern recognition. The inclusion measure of fuzzy sets represents the degree to which a fuzzy set is contained in another fuzzy set. Sinha [8] first introduced an axiomatic definition of the inclusion measure of fuzzy sets. Later on, Cornelis [4] revised sinha's axiom, and proposed a inclusion for intuitionistic fuzzy set. Bustince [2] investigated the inclusion grade for interval-valued fuzzy sets.

Information entropy, as a measure of fuzziness often used in the literature, describe the fuzziness degree or uncertain information of a fuzzy set and is first mentioned by Zadeh. It has received great attention recently. In 1972 [5] Deluca and Termini presented some axioms to describe the fuzziness degree of fuzzy set, with which a fuzzy entropy based on Shannon's function was proposed. After that, many other researchers have studied the fuzzy entropy in different ways. Especially, Shang in [9] introduced another new fuzziness measure. For the information measure of intuitionistic fuzzy set, Szmidt [10] proposed an entropy for intuitionistic fuzzy set by employing a geometric interpretation of IFS. And Vlachos [11] studied the intuitionistic fuzzy information entropy and its application to pattern recognition. In 2010 Ye [15] also proposed two effective measures of intuitionistic fuzzy entropy. Moreover, Burillo [3] presented an entropy on interval-valued fuzzy sets and on intuitionistic fuzzy sets. And Vlachos [12] studied the subsethood and entropy of interval-valued fuzzy sets. Additionally, we presented some entropy formulae of interval-valued intuitionistic fuzzy sets [16,17].

However, by now, although some information entropy formulae, similarity measures and inclusion measures of IFSs have been introduced, and applied to many real life problems, there is little investigation on their relationships among them, which may be widely applied to many fields such as pattern recognition, cluster analysis, image segment. Therefore, it is worthwhile to focus on discovering the close relationships among entropy, similarity measure and inclusion measure for IFS. In section 2, we first summarize the axiomatic definitions of entropy, similarity measure and inclusion measure for IFS. Further, in section 3, some theorems reflecting the relationship among the entropy, the similarity measure, the inclusion measure are obtained, based on which some new formulae to evaluate information entropy and similarity, inclusion measure of IFS are then put forward.

2 Preliminaries

Definition 2.1 [1]. An intuitionistic fuzzy set (IFS) A in the finite universe $X = \{x_1, x_2, \dots, x_n\}$ is a mapping $A: X \rightarrow [0,1] \times [0,1]$, i.e., $A(x_i) = (u_A(x_i), v_A(x_i))$ satisfying $0 \leq u_A(x) + v_A(x) \leq 1$, $\forall x_i \in X$, where $u_A(x_i)$, $v_A(x_i)$ represent the degree of membership and nonmembership of that element x_i to the set A , respectively. We denote by $IF(X)$ all the IFSs in X .

Definition 2.2. Let $A, B \in IF(X)$, some basic operations are defined as follows[13]:

$$A \cup B = \{< x_i, (u_A(x_i) \vee u_B(x_i), v_A(x_i) \wedge v_B(x_i)) \geq x_i \in X\};$$

$$A \cap B = \{< x_i, (u_A(x_i) \wedge u_B(x_i), v_A(x_i) \vee v_B(x_i)) \geq x_i \in X\};$$

$$A' = \{< x_i, (v_A(x_i), u_A(x_i)) \geq x_i \in X\};$$

$$A \subseteq B, \text{if } [u_A(x_i) \leq u_B(x_i), v_A(x_i) \geq v_B(x_i)], \quad \forall x_i \in X;$$

$$A = B, \text{if } u_A(x_i) = u_B(x_i), v_A(x_i) = v_B(x_i), \forall x_i \in X;$$

Definition 2.3. A real function $E : IF(X) \rightarrow [0,1]$ is named an information entropy of IFSs on finite universe X, if E satisfies all the following properties[10]:

- (P1) $E(A) = 0$ if A is a crisp set, $\forall A \in P(X)$;
- (P2) $E(A) = 1$ if $u_A(x_i) = v_A(x_i), \forall v_A(x_i) \in X$;
- (P3) $E(A) \leq E(B)$, if A is less fuzzy than B ($A \ll B$), which is defined as
 $u_A(x_i) \leq u_B(x_i), v_A(x_i) \geq v_B(x_i)$, for $u_B(x_i) \leq v_B(x_i)$;
 $u_A(x_i) \geq u_B(x_i), v_A(x_i) \leq v_B(x_i)$, for $u_B(x_i) \geq v_B(x_i)$;
- (P4) $E(A) = E(A')$.

Below, we display two entropy formulae of intuitionistic fuzzy set A , which fulfill all the above conditions (P1 – P4).

$$E_1(A) = 1 - \frac{1}{n} \sum_{i=1}^n |u_A(x_i) - v_A(x_i)| ;$$

$$E_2(A) = 1 - \sqrt{\frac{1}{n} \sum_{i=1}^n |u_A(x_i) - v_A(x_i)|^2} .$$

Definition 2.4. A real function $S : IF(X) \times IF(X) \rightarrow [0,1]$ is named as the similarity measure of IFSs on universe X, if it satisfies properties [7]:

- (S1) $S(A, A') = 0$, if A is a crisp set, $\forall A \in p(X)$;
- (S2) $S(A, B) = 1$, if $A = B$;
- (S3) $S(A, B) = S(B, A)$;
- (S4) If $A \subseteq B \subseteq C$, then $S(A, C) \leq S(A, B)$ and $S(A, C) \leq S(B, C)$.

For example, the following similarity measures of IFSs had been proposed:

$$S_1(A, B) = 1 - \frac{1}{2n} \sum_{i=1}^n |u_A(x_i) - u_B(x_i)| + |v_A(x_i) - v_B(x_i)| ;$$

$$S_2(A, B) = 1 - \sqrt{\frac{1}{2n} \sum_{i=1}^n |u_A(x_i) - u_B(x_i)|^2 + |v_A(x_i) - v_B(x_i)|^2} .$$

Definition 2.5. A real function $I : IF(X) \times IF(X) \rightarrow [0,1]$ is named as the inclusion measure of IFSs on universe X, if it satisfies the following properties:

- (I1) $I(X, \varphi) = 0$;
- (I2) $I(A, B) = 1$ if $A \subseteq B$;
- (I3) if $A \subseteq B \subseteq C$, then $I(C, A) \leq I(B, A)$ and $I(C, A) \leq I(C, B)$.

Obviously, if the IFSs A, B become fuzzy sets, then $I(A, B)$ reduces to the subsethood of fuzzy sets. Now, we give two inclusion measures between IFSs A, B as below.

$$I_1(A, B) = 1 - \frac{1}{2n} \sum_{i=1}^n \{|u_A(x_i) - u_A(x_i) \wedge u_B(x_i)| + |v_A(x_i) \vee v_B(x_i) - v_A(x_i)|\};$$

$$I_2(A, B) = 1 - \sqrt{\frac{1}{2n} \sum_{i=1}^n \{|u_A(x_i) - u_A(x_i) \wedge u_B(x_i)|^2 + |v_A(x_i) \vee v_B(x_i) - v_A(x_i)|^2\}}$$

3 Relationships among Information Entropy, Similarity Measure, Inclusion Measure of IFSs

According to the previous discussion, one notice that the real functions of information entropy, similarity measure and inclusion measure of IFSs are not unique. Therefore, in what follows we will investigate on some important relationships among the similarity measure, the distance measure, the inclusion measure and the entropy of IFSs in detail. And then we will put forward some novel formulae to calculate the similarity measure, Inclusion measure and entropy of IFSs.

Theorem 3.1. Assume S is a similarity measure of IFSs, for $A, B \in IF(X)$, then $d(A, B) = 1 - S(A, B)$ is a distance measure between IFSs A and B .

Proof. The proof is straightforward (omitted).

Theorem 3.2. Assume S is a similarity measure of IFSs, let $A \in IF(X)$, then $E(A) = S(A, A')$ is an information entropy of IFS A .

Proof. We only need to prove that all the properties in Def 2.3 hold.

(P1): If A is a crisp set, i.e., $A \in p(x)$, and due to that S is a similarity measure of IFSs, then from definition 2.4, we have $S(A, A') = 0$, Thus $E(A) = 0$.

(P2): If $u_A(x_i) = v_A(x_i)$, $\forall x_i \in X$, then we know

$u_{A'}(x_i) = v_A(x_i) = u_A(x_i)$, $v_{A'}(x_i) = u_A(x_i) = v_A(x_i)$. Hence, $A' = A$. Thus, by the Definition 2.4 of similarity measure, we get $E(A) = S(A, A') = 1$

(P3): If A is less fuzzy than B , denoted by $A \ll B$, then we know, when $u_A(x_i) \leq u_B(x_i)$, $v_A(x_i) \geq v_B(x_i)$, for $u_B(x_i) \leq v_B(x_i)$;

i.e., $u_A(x_i) \leq u_B(x_i) \leq v_B(x_i) \leq v_A(x_i)$, so, $A \subseteq B \subseteq B' \subseteq A'$.

Therefore, by the definition 2.4 of similarity measure of IFSs, we have

$$S(A, A') \leq S(B, A') \leq S(B, B')$$

Similarly, when $u_A(x_i) \geq u_B(x_i)$, $v_A(x_i) \leq v_B(x_i)$, for $u_B(x_i) \geq v_B(x_i)$;

i.e., $u_A(x_i) \geq u_B(x_i) \geq v_B(x_i) \geq v_A(x_i)$,

So, $A \supseteq B \supseteq B' \supseteq A'$.

Therefore, by the definition of similarity measure of IFSs, we get

$$E(A) = S(A, A') \leq S(B, A') \leq S(B, B') = E(B)$$

(P4): From the definition of similarity measure of IFSs, we know that

$$S(A', A) = S(A, A') , \text{ for any } A \in IF(X) .$$

Hence $E(A') = E(A)$.

Theorem 3.3. Suppose S is a similarity measure of fuzzy sets, assume $A = \langle u_A, v_A \rangle \in IF(X)$, $X = \{x_1, x_2, \dots, x_n\}$, and u_A, v_A are the membership and non-membership degree of IFS A , respectively, then $S(u_A, v_A)$ is an entropy of IFS A .

Proof

(P1): If $A \in p(X)$, then for any $x_i \in X$,

$$u_A(x_i) = 1, v_A(x_i) = 0 \text{ or } u_A(x_i) = 0, v_A(x_i) = 1 ;$$

$$\text{i.e., } v_A(x_i) = (u_A(x_i))', \forall x_i \in X$$

Since S is the similarity measure of fuzzy sets, we have

$$S(u_A, v_A) = S(u_A, (u_A)') = 0 .$$

(P2): From the definition of similarity measure, we immediately have

$$S(u_A, v_A) = 1 \Leftrightarrow u_A = v_A, \text{ i.e. } u_A(x_i) = v_A(x_i), \forall x_i \in X .$$

(P3): Since A is less fuzzy than B , then we know,

when $u_A(x_i) \leq u_B(x_i)$, $v_A(x_i) \geq v_B(x_i)$, for $u_B(x_i) \leq v_B(x_i)$;

i.e., $u_A(x_i) \leq u_B(x_i) \leq v_B(x_i) \leq v_A(x_i)$,

So, $u_A \subseteq u_B \subseteq v_B \subseteq v_A$.

Known by the definition of similarity measure, we have

$$S(u_A, v_A) \leq S(u_A, v_B) \leq S(u_B, v_B) .$$

When $u_A(x_i) \geq u_B(x_i)$, $v_A(x_i) \leq v_B(x_i)$, for $u_B(x_i) \geq v_B(x_i)$;

then, $u_A(x_i) \geq u_B(x_i) \geq v_B(x_i) \geq v_A(x_i)$,

So, $u_A \supseteq u_B \supseteq v_B \supseteq v_A$.

By the definition of similarity measure, we get

$$E(A) = S(u_A, v_A) \leq S(u_B, v_A) \leq S(u_B, v_B) = E(B) .$$

(P4): Since $A' = \langle v_A, u_A \rangle$ and from the definition of similarity, it follows that

$$E(A') = S(v_A, u_A) = S(u_A, v_A) = E(A) .$$

Theorem 3.4. Let E be an entropy of IFS, for any two IFSs $A, B \in IF(X)$, we define another intuitionistic fuzzy set $h(A, B) \in IF(X)$ as follows:

$$u_h(A, B)(x) = \frac{1 + [\frac{1}{2}(|u_A(x) - u_{A \cap B}(x)| + |v_A(x) - v_{A \cap B}(x)|)]^2}{2} ,$$

$$v_h(A, B)(x) = \frac{1 - [\frac{1}{2}(|u_A(x) - u_{A \cap B}(x)| + |v_A(x) - v_{A \cap B}(x)|)]}{2} ,$$

then $E(h(A, B))$ is an inclusion measures of IFSs A and B .

Proof

(I1) If $A = X$, $B = \emptyset$, then

$$u_h(A, B)(x) = (u_h(A, B)(x), v_h(A, B)(x)) = (1, 0), \quad \forall x \in X . \text{ i.e., } h(A, B) \in P(X) ;$$

Again by axiomatic property of entropy we know that

$$E(h(A, B)) = 0.$$

(I2) If $A \subseteq B$, then $u_h(A, B) = v_h(A, B) = 1/2$. Thus $E(h(A, B)) = 1$.

(I3) If $A \subseteq B \subseteq C$, then $u_h(C, A) \geq u_h(C, B), v_h(C, A) \leq v_h(C, B)$,

Again from $u_h(C, B) \geq v_h(C, B)$, we know that $h(C, A) \ll h(C, B)$.

Thus, by the axiomatic definition of entropy we have $E(h(C, A)) \leq E(h(C, B))$.

Similarly, we can get $E(h(C, A)) \leq E(h(B, A))$.

Theorem 3.5. Let I be an inclusion measure of IFSs, for each $A \in \text{IF}(X)$, A' be the complement of IFS A , then $E(A) = I(A \cup A', A \cap A')$ is an entropy of IFS A . Proof. We need to prove that it satisfies all the axiomatic properties of entropy for IFS.

(P1): If $A \in P(X)$, then $A \cup A' = X, A \cap A' = \emptyset$,

$$\text{So, } I(A \cup A', A \cap A') = I(X, \emptyset).$$

Since I is an inclusion measure, we have $I(A \cup A', A \cap A') = I(X, \emptyset) = 0$.

(P2): From property **(I2)** in the definition of inclusion measure, we immediately have

$$I(A \cup A', A \cap A') = 1 \Leftrightarrow A \cup A' = A \cap A'$$

$$\Leftrightarrow A = A' \Leftrightarrow u_A(x_i) = v_A(x_i), \forall x \in X.$$

(P3): If A is less fuzzy than B , then we easily deduce that

$$A \cap A' \subseteq B \cap B' \subseteq B \cup B' \subseteq A \cup A'.$$

With the property **(I3)** in definition 2.5 of inclusion measure, we get

$$E(A) = I(A \cup A', A \cap A')$$

$$\leq I(B \cup B', A \cap A')$$

$$\leq I(B \cup B', B \cap B') = E(B).$$

(P4): Since $A' = \langle v_A, u_A \rangle, (A')' = A$, we have

$$E(A') = I(A' \cup A, A' \cap A) = E(A).$$

4 Conclusions

In this paper, we present some new axioms and formulae of computing information entropy measures, similarity measures and inclusion measures, which have been applied to many real-world fields such as pattern recognition, clustering analysis and image processing. The close relationships among information entropy, similarity measure, inclusion measure of IFSs are also investigated in detail.

Acknowledgments. This work is supported by the Humanities and Social Sciences Youth Foundation of Ministry of Education of China (No. 12YJCZH281), the Guangzhou Social Science Planning Project. “The study of early warning index selection and urgent decision mechanism for city significant emergency in uncertain

environment" (No. 2012GJ31), the National Natural Science Foundation (Nos. 60974019, 61070061, 60964005), and the Guangdong Province High-level Talents Project.

References

1. Atanassov, K.: Intuitionistic fuzzy sets. *Fuzzy Sets and Systems* 20, 87–96 (1986)
2. Bustince, H.: Indicator of inclusion grade for interval-valued fuzzy sets: Application to approximate reasoning based on interval-valued fuzzy sets. *International Journal of Approximate Reasoning* 23, 137–209 (2000)
3. Burillo, P., Bustince, H.: Entropy on intuitionistic fuzzy sets and on interval-valued fuzzy Sets. *Fuzzy Sets and Systems* 78, 305–316 (1996)
4. Cornelis, C., Kerre, E.E.: Inclusion Measures in Intuitionistic Fuzzy Set Theory. In: Nielsen, T.D., Zhang, N.L. (eds.) ECSQARU 2003. LNCS (LNAI), vol. 2711, pp. 345–356. Springer, Heidelberg (2003)
5. DeLuca, A., Termini, S.: A definition of nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control* 20, 301–312 (1972)
6. Hung, W.L.: Similarity measures of intuitionistic fuzzy sets based on Hausdorff distance. *Pattern Recognition Letters* 25, 1603–1611 (2004)
7. Szmidt, E., Kacprzyk, J.: A Similarity Measure for Intuitionistic Fuzzy Sets and Its Application in Supporting Medical Diagnostic Reasoning. In: Rutkowski, L., Siekmann, J.H., Tadeusiewicz, R., Zadeh, L.A. (eds.) ICAISC 2004. LNCS (LNAI), vol. 3070, pp. 388–393. Springer, Heidelberg (2004)
8. Sinha, D., Dougherty, E.R.: Fuzzification of set inclusion: theory and applications. *Fuzzy Sets and Systems* 55, 15–42 (1993)
9. Shang, X.G., Jiang, W.S.: A note on fuzzy information measures. *Pattern Recognition Letters* 18, 425–432 (1997)
10. Szmidt, E.: Entropy for intuitionistic fuzzy sets. *Fuzzy Sets and Systems* 118, 467–477 (2001)
11. Vlachos, I.K.: Intuitionistic fuzzy information-Applications to pattern recognition. *Pattern Recognition Letters* 28, 197–206 (2007)
12. Vlachos, I.K.: Subsethood, entropy, and cardinality for interval-valued fuzzy sets - An algebraic derivation. *Fuzzy Sets and Systems* 158, 1384–1396 (2007)
13. Wei, G.W.: GRA method for multiple attribute decision making with incomplete weight information in intuitionistic fuzzy setting. *Knowledge-Based Systems* 23, 243–247 (2010)
14. Ye, J.: Cosine similarity measures for intuitionistic fuzzy sets and their applications. *Mathematical and Computer Modelling* 53, 91–97 (2011)
15. Ye, J.: Two effective measures of intuitionistic fuzzy entropy. *Computing* 87, 55–62 (2010)
16. Zhang, Q.S., Jiang, S.Y., Jia, B.G., Luo, S.H.: Some information entropy measures for interval-valued intuitionistic fuzzy sets. *Information Sciences* 180, 5130–5145 (2010)
17. Zhang, Q.S., Jiang, S.Y.: Relationships between Entropy and Similarity Measure of Interval-Valued Intuitionistic Fuzzy Sets. *International Journal of Intelligent Systems* 25, 1121–1140 (2010)

Optimization of Tandem Cold Rolling Schedule Based on Collaborative Optimized PSO

Li Ying, Wang Jing-sheng, Wang Hong-rui, and Wei Li-xin

Key Lab of Industrial Computer Control Engineering of Hebei Province,

Yanshan University, Qinhuangdao 066004, China

wlx2000@ysu.edu.cn, wangjingsheng2005@163.com

Abstract. The reasonable rolling schedule is not only beneficial to improve the accuracy and achieve good shape of cold rolled steel strip, but also has practical value in prolonging the service life of equipment and improving the production efficiency of enterprise. It tends to reach premature convergence when particle swarm optimization algorithm is applied in the optimization of rolling schedule. Based on the rapid convergence of particle swarm optimization algorithm and evenly traversal of Tent sequence, a collaborative optimization algorithm which combines particle swarm optimization algorithm with chaos searching is introduced in this paper. The proposed algorithm can overcome the disadvantage that particle swarm optimization algorithm easily falls into the local minimum, and find pressure rate satisfying the preset target function using less iteration times and optimal time, and realize the optimal rate target.

Keywords: Particle Swarm Optimization, Chaos searching, Collaborative optimize, Rolling schedule optimization.

1 Introduction

Rolling schedule of tandem cold rolling mill must set reductions, speeds, tensions etc. according to the thickness and width of incoming materials, the thickness of finished strip, the type of steel and the requirements of equipment and technology. For tandem cold rolling mill of a strong coupled mechanical structure, the rolling parameters of all stands will be changed when one stand is changed. So it is particularly important to use suitable arithmetic to formulate and optimize rolling schedule. The traditional methods, such as the energy consume method and schedule method, mainly depended on experience, are simple to implement but lack of flexibility and accuracy, thus can not meet the needs of modern production. In recent years, researchers apply approaches like Newton-Raphson[1], Data Mining[2], and Momentum Technique[3], etc. to load distribution, and acquire good effect. In Ref. [4-6], genetic algorithm and adaptive genetic algorithm are applied in single-objective and multi-objective optimization of tandem cold rolling schedule to improve the seeking accuracy and successful rate. However, the weakness of slow optimal speed of genetic algorithm is also displayed, it is unfavorable for optimization of tandem cold rolling schedule. In Ref. [7], PSO, which has fast convergence speed, is used in optimization of tandem

cold rolling schedule, but there is still some disadvantages of local minimum and premature convergence. For the mentioned problem, this paper proposes the collaborative optimized PSO which reduces the iteration times and searching time on the premise of improvement of global searching ability by combining the PSO algorithm and chaos searching together through mutual information sharing.

2 Collaborative Optimization PSO

2.1 Particle Swarm Optimization Algorithm and Optimization

Particle Swarm Optimization(PSO) algorithm[8], proposed by Kennedy and Eberhart in 1995, the velocity and the position of a particle are updated by

$$\begin{cases} v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_i - x_i(t)) + c_2 r_2 (P_g - x_i(t)) \\ x_i(t+1) = x_i(t) + v_i(t+1) \end{cases} \quad (1)$$

where $i = 1, 2, \dots, n$, n -the size of swarm. c_1, c_2 -nonnegative constants called acceleration coefficients, r_1, r_2 - random numbers within $[0, 1]$, ω -an inertia weight to control particle's exploration capability in the search space. p_i - the current optimal value of the i -th particle, P_g -the current optimal value of the swarm.

In the speed updating formula, $c_1 r_1 (p_i - x_i(t))$ can be viewed as cognitive components, quantify dependence on the previous experience. $c_2 r_2 (P_g - x_i(t))$ can be regarded as social components, its function is to pull the search direction to the current optimum locations of its neighbors. In collaborative optimization algorithm, the task of PSO is to find out global optimal points rapidly to reflect the rapidity of PSO, so the cognitive components are removed, which is rewritten as

$$\begin{cases} v_i(t+1) = \omega v_i(t) + c_2 r_2 (P_g - x_i(t)) \\ x_i(t+1) = x_i(t) + v_i(t+1) \end{cases} \quad (2)$$

PSO only with social composition has the better rapidity than the original, which has been clarified empirically by Kennedy [9,10] .

2.2 Chaos Searching

Chaos searching uses chaotic sequences to produce feasible solutions in the searching space, the ability to jump out of local minima has been enhanced by the process of chaos sequences traversal. Tent map is selected in the collaborative optimization PSO, it is proposed that the traverses uniformity of Tent map is better than the Logistic map. In Ref. [11], the Tent Mapping is defined as follow:

$$L(k+1) = \begin{cases} 2L(k), & 0 \leq L(k) \leq 1/2 \\ 2(1-L(k)), & 1/2 < L(k) \leq 1 \end{cases} \quad (3)$$

However, there are still some shortages: firstly, Tent Map has small cycle points and unstable periodic points, such as 4 cycle number (0.2, 0.4, 0.8, 0.6) and 0.25, 0.5, 0.75, all of them iterate towards the fixed point 0. Secondly, the operations of internal computer is to use binary operations, while the Tent Map is also a operation multiplied by 2, so the sequence values will tend to 0 after a number of iterations.

Considering the above two aspects, an average value of 7 random numbers is regarded as $L(0)$ at beginning of the sequence initialization. Because 7 is a prime, it will not easily produce small cycle points and unstable periodic points mentioned above. Particles are distributed to the current searching space uniformly by the Tent map according to equation as follow:

$$x(i, j) = x_{\min_d}(j) + (x_{\max_d}(j) - x_{\min_d}(j)) \times L(i) \quad (4)$$

where $L(i)$ - i -th Chaos variable produced by Eq.(3), $[x_{\min_d}(j), x_{\max_d}(j)]$ - j -th dimension of current searching space. So the iterations of Tent map is the number of particles. If the accuracy of the data is set to 10^{-14} , the value of Tent map doesn't tend to 0. So the problems existing in Tent map are resolved very well.

2.3 Realization of Collaborative Optimization PSO

Usually, chaos searching is a vertical conduct[12] in hybrid optimization algorithm. When PSO occurs early stagnation or trap in local minimum, chaos searching can make it jump out of the local minimum or avoid early stagnation[13], however, the disadvantage is that the searching time is greatly increased. The collaborative optimization PSO is different in this paper. The two search always run through the entire optimization process. PSO is responsible for global searching, making particles gathered quickly and achieving rapidity. Chaos searching looks for optimization points or the adjacent points through traversing the whole searching space uniformly, then pass on these points to PSO and also make it jump out of the local minimum and avoid early stagnation.

Firstly, all particles are divided into two swarms which are named swarm P for particle swarm optimization and swarm C for chaos searching. Secondly, update the global optimization point P_g after each searching and the parameters of swarm P for the next PSO, at the same time, Chaos searching space is redefined as follow:

$$\begin{cases} x_{\max_d}(j) = P_g(j) + (x_{\max}(j) - x_{\min}(j)) \times 0.5^t; \\ x_{\min_d}(j) = P_g(j) - (x_{\max}(j) - x_{\min}(j)) \times 0.5^t; \end{cases} \quad (5)$$

and whose range is limited. Specific algorithm is shown as Fig. 1, where N_P , N_C -the particles number of swarm P and swarm C, N-the number of all particles, $N=N_P+N_C$. Inertia weight in the algorithm is used a linear diminishing strategy proposed by Eberhart and Shi[14].

3 Object Functions and Rolling Model

3.1 Determinate Object Functions and Constraints

Object functions of tandem cold rolling schedule is various based on the differences of production needs, equipment factors and technology factors. The equal relatively rolling power is adopted as object function in this paper. Specific object function is shown as

$$\min J = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{P_i}{P_{Ni}} - \frac{P_j}{P_{Nj}} \right) \quad (6)$$

where P_i, P_j -rolling power of i -th, j -th stand, P_{Ni}, P_{Nj} - the rated power of main motor of i -th, j -th stand, J -object function value, n -number of stands.

Constraints can be divided into two categories which are the technology constraints and equipment constraints. Technology constraints include $\varepsilon_{min} \leq \varepsilon \leq \varepsilon_{max}$, $n_{min} \leq n \leq n_{max}$, $T_{min} \leq T \leq T_{max}$, where ε , n , T -the actual rolling reduction, roller speed and tension, ε_{min} , ε_{max} , n_{min} , n_{max} , T_{min} , T_{max} -minimum and maximum of rolling reduction, roller speed and tension. Device constraint include $F_i \leq F_{max}$, $M_i \leq M_{max}$, $P_i \leq P_{max}$, where F_i , M_i , P_i -the actual rolling force, rolling torque and rolling power, F_{max} , M_{max} , P_{max} -maximums of rolling force, rolling torque and rolling power.

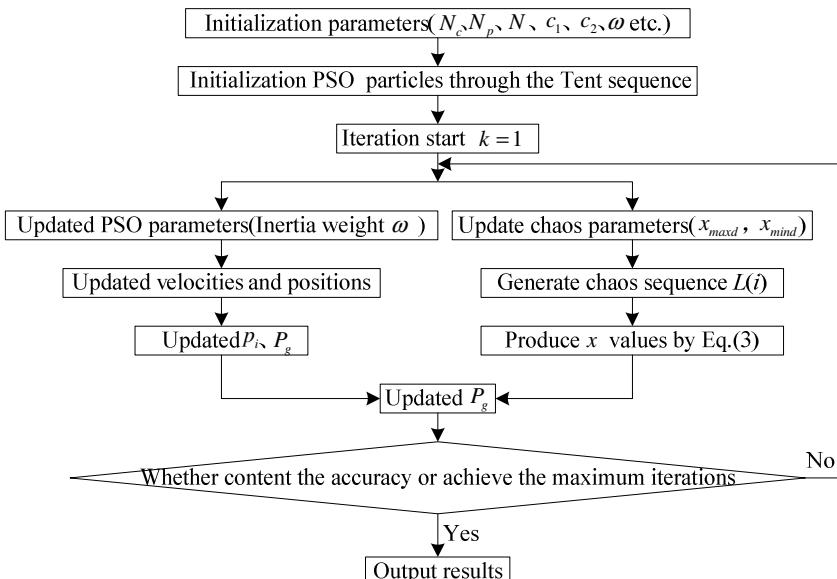


Fig. 1. The flow chart of Collaborative optimization algorithm

3.2 Rolling Force Models

The Bland-Ford-Hill implicit model is adopted as rolling force model, which is given by

$$F = Bl_c' Q_p K_T K / 1000 \quad (7)$$

where F -rolling force, kN, B -width of incoming materials, mm, K -resistance to deformation considering the principal stress influence coefficients in width direction, usually taking $K = 1.15\sigma$ when lateral spread is small, MPa, σ - material deformation resistance, MPa, K_T -tension influence coefficient, $K_T = 1 - (0.7\tau_b + 0.3\tau_f) / K$, where τ_f - forward tension, τ_b -backward tension, $\tau_b, \tau_f = (0.2 \sim 0.4)K$, l_c' - length of contact arc considering contact area of flattening zone, mm, Q_p -the influence of friction coefficient $Q_p = 1.08 + 1.79\mu\varepsilon\sqrt{1-\varepsilon}\sqrt{R'/h} - 1.02\varepsilon$, where μ -friction coefficient, ε -the relative rolling reduction, $\varepsilon = \Delta h / h_m$.

3.3 Rolling Power Model

Rolling power model[5] is descript as follow:

$$P = M\omega \quad (8)$$

where M -rolling torque, Nm, $M = 2F\psi l_c'$, F -rolling force, kN, ψ - coefficient of force arm, $\psi = 0.6489 - 0.8066\varepsilon + 0.3491\varepsilon^2$, ω - angular velocity of roller, rad/s.

4 Design of Rolling Schedule Optimization

Taking into account the actual production, equal relatively rolling power distribution is only used in the first four stands, the last stand is usually used as a pinch pass mill, adopt a fixed reduction as follow:

$$\varepsilon_5 = (h_4 - h_5) / h_4 \quad (9)$$

where ε_5 - reduction of the last stand, which is adopted 10%~15% because it is easy to appear the phenomenon of velocity inversion when the reduction rate is less than 10%, h_5 - exit thickness of the last stand, mm, h_4 - entry thickness of the last stand, mm. It selects the reduction rate distribution ratio $E = (e_1, e_2, e_3, e_4)$ as the independent variable of object function Eq.(6), the reduction rate distribution ratio can not be greater than 45% on the basis of experience, so it takes $e_i \in [10, 45], i = 1 \sim 4$. And

then the reduction rate distribution ratio is converted to the reduction rate, which can get the entry thickness and exit thickness on each stand after that. The object function value can be gotten by means of applying the above data to Eq.(6). At the same time $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ is taken as optimal particles of collaborative optimized PSO. Where $x_i = e_i, i = 1 \sim 4; x_5 = \varepsilon_5 \times 100$ and the region of \mathbf{x} is $\{x | x_i \in [10, 45], i = 1 \sim 4; x_5 \in [10, 15]\}$. The specific optimization procedures are in Fig.2.

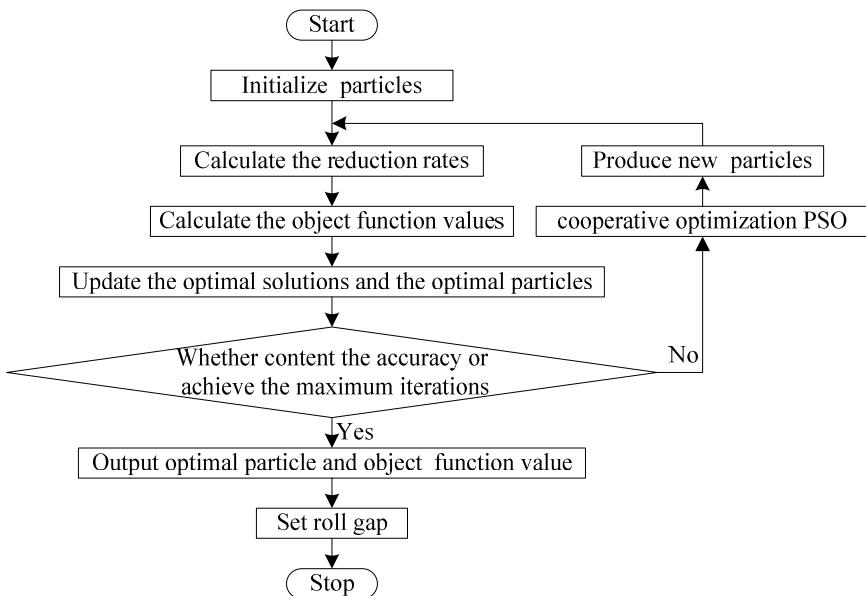


Fig. 2. The flow chart of schedule optimization

5 Instance Validation

Taking a 1250 8-high 5-stand cold tandem rolling of a certain rolling mills in Jiangsu Province for example. Equipment parameters are shown in table 1.

Table 1. The parameters of 1250 8-high 5-stand cold tandem rolling

	Stands				
	1	2	3	4	5
Motor rated power (kW)	1300x2	1300x3	1300x3	1300x3	1300x3
Motor speed (r/min)	325/800	325/800	325/800	325/800	325/800
Maximum of rolling force (t)	1100	1100	1100	1100	1100
Radius of work roll (mm)	310~290	280~265	290~270	300~280	310~290

Type of steel is St14, and deformation resistance formula as follow:

$$\sigma_s = 282.25 + (89497.28 \times \varepsilon' + 103.61)^{0.5491} \quad (10)$$

where ε' -cumulative deformation degree, the size of incoming materials is 3.0mm×1000mm, the size of product is 0.4mm×1000mm, highest rolling speed is 16m/s.

Cooperative optimization PSO and the method presented in Ref. [7] are applied into rolling schedule optimization of 1250 8-high 5-stand cold tandem rolling, maximum iteration number is set to 200, 50 experiments are taken, $N_p=N_c=20$. It is considered successful when the object value is less than 10^{-20} . Results are as follows:

Table 2. Comparisons for the simulation results of the schedule optimization

	The success numbers	Average success iterations
Method in Ref. [7]	33	120.93
Cooperative optimization PSO	50	74.78

To analysis further, it selects two successful optimization curves to compare as follow:

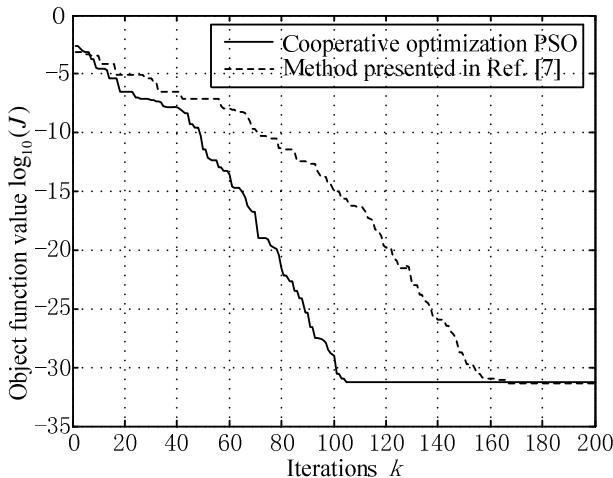


Fig. 3. The contrast for the curves of schedule optimization

Table 3. Specification for optimal results

Stands	1	2	3	4	5
Exit thickness (mm)	2.17	1.23	0.74	0.45	0.40
Rolling force (t)	824.6	838.4	666.7	530.4	285.6
Rolling power (kW)	823.483	1235.224	1235.224	1235.224	495.275

Optimal results gotten by collaborative optimization PSO is shown as Table 3. From the simulation results we can see that, on the one hand, collaborative optimization PSO using in the rolling schedule makes the optimizing success rate and the convergence speed enhance obviously. On the other hand, chaos searching uses less time than particle swarm optimization.

6 Conclusion

This paper uses particle swarm optimization and collaborative optimization PSO to optimize the rolling schedule of 1250 8-high 5-stand cold tandem rolling, which takes equal relatively rolling power as object function. The simulation results show that collaborative optimization PSO is better than particle swarm optimization in terms of optimization ability, which can jump out of the local minimum points to reach a better precision, and it is also better than particle swarm optimization not only in the iteration times but also in the searching time. From simulation results on the success rates and iterations, a conclusion can be drawn that the rolling schedule optimization process using collaborative optimization PSO is completed more quickly and can seek results more accurate.

References

1. Pires, C.T.A., Ferreira, H.C., Sales, R.M., et al.: Set-up Optimization for Tandem Cold Mills: a Case Study. *Journal of Materials Processing Technology* 173(3), 368–375 (2006)
2. Zhang, Y., Liu, X., Wang, G.: Research on Plate Rolling Load Distribution Based on Data Mining. *Iron and Steel* 40(4), 44–45 (2005)
3. Li, H., Xu, J., Gong, D., et al.: Application of Momentum Technique in Load Distribution for Tandem Hot Strip Mill. *Iron and Steel* 41(2), 46–50 (2006)
4. Li, Y., Liu, J., Wang, Y.: An improved adaptive weight approach GA for optimizing multi-objective rolling schedules in a tandem cold rolling. *Control Theory and Applications* 26(6), 687–693 (2009)
5. Wei, L., Li, X., Li, Y., et al.: Optimization of Tandem Cold Rolling Schedule Based on Improved Adaptive Genetic Algorithm. *Journal of Mechanical Engineering* 46(16), 136–141 (2010)
6. Yang, J., Dou, F., Liu, S., et al.: Application of genetic algorithm to rolling schedule in tandem cold mill. *China Mechanical Engineering* 18(15), 1868–1871 (2007)
7. Li, Y., Liu, J.C., Wang, Y.: An Adaptive Weight PSO for Rolling Schedules Multi-objective Optimization of Tandem Cold Rolling. In: Proceedings of the IEEE International Conference on Automation and Logistics, vol. 8, pp. 895–899 (2009)
8. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proc. of IEEE Int. Conf. on Piscataway, pp. 1942–1948 (1995)
9. Kennedy, J.: The Particle Swarm: Social Adaptation of Knowledge. In: Proceedings of the IEEE International Conference on Evolutionary Computation, vol. 4, pp. 303–308 (1997)
10. Hong, T., Peng, G., Li, Z.P., Liang, Y.: A Novel Evolutionary Strategy for Particle Swarm Optimization. *Chinese Journal of Electronics* 18(4), 771–774 (2009)

11. Shan, L., Qiang, H.: Chaotic optimization algorithm based on Tent map. *Control and Decision* 2, 179–182 (2005)
12. Zhang, X., Wen, S., Li, H.: Chaotic Particle Swarm Optimization Algorithm Based on Tent Mapping. *China Mechanical Engineering* 19(17), 2108–2112 (2008)
13. Meng, H., Zheng, P.: Particle Swarm Optimization Algorithm Based on Chaotic Series. *Control and Decision* 3, 263–266 (2006)
14. Eberhart, R.C., Shi, Y.: Particle swarm optimization. In: Proc. of Congress on Evolutionary Computation, pp. 81–88 (2001)

Career Decision-Making of University Students Based on Analytic Hierarchy Process

Ming Lu

Department of Information Engineering, Jilin Business and Technology College 130062,
Changchun, China
lum238@sina.com

Abstract. This paper use analytic hierarchy process (AHP). According to the results of the questionnaire obtained six main factors affecting the choice of work, According to Undergraduate career choice of AHP model, by constructing judgment matrix, carry out hierarchical sorting and consistency test, Assignment for each impact factors, according to the assignment the right to re-sort, analyze and identify the main factors that influence graduates career. Provide reference for university students in choosing the concept of education and employment guidance. Through the analysis and solution of the subject, we look to see that AHP method advantage of systematic, practical, and the simplicity.

Keywords: AHP, Judgment Matrix, Consistency Test, Structural Model, College Students' Career.

1 Introduction

With the successive year's rapid growth in the number of university graduates, the graduates' employment problem has become more apparent. Correct and effective employment guidance for the graduates, and should be established in the understanding of students' employment mentality, especially to understand the affect their career satisfaction factors and the degree of their importance in the minds of students. The factors that affect students' career satisfaction are multifaceted, and complex. Through research that these factors are: the prospects for development; interpersonal relationships; wages; colleagues circumstances; unit known; the mutual ; value orientation; family concept, and so on relationship between these factors of location[1]. This paper aims to conduct a survey of university students about to go to work and understand their selection of the employer to the major consideration, and ultimately through the AHP to determine the weight order of the jobs, so as to provide a reliable reference for University Graduates.

2 Analytic Hierarchy Process

Analytic hierarchy process (AHP) is puts forward by the United States operations research scientist at the University of Pittsburgh Professor Saaty in the 1970s. for the

United States Department of defense research" according to various industrial sectors of national welfare contribution size and power allocation" task, which a analytic hierarchy analysis method.

The Analytic Hierarchy Process (AHP) [2] is a qualitative and quantitative combination of multiple criteria decision making method, it can subjective judgment of the person's use number forms of expression and processing. Analytic hierarchy process is a complicated problem into a number of factors to identify the relationship between that contain these factors, and the level of these elements, in accordance with the scheme layer, object layer, rule layer arranged together to form a ladder hierarchy model, In each layer according to certain criteria, the various elements of the layer compared to the establishment of judgment matrix. For each judgment matrix, the application of traditional feature vector method, find the corresponding eigenvectors, obtained priority vectors of the layer on layer of an element; Then, according to the principle of analytic hierarchy process to calculate the combination weight of layers elements for the overall objective, to arrive at the final weights of the different scenarios and provide the basis for the selection of the optimal solution [3].

3 Model of College Students' Career Based on AHP

The use of analytic hierarchy process modeling, generally in accordance with the following four steps: (1) the establishment of hierarchical structure model; (2) constructed at all levels of all judgment matrix; (3) hierarchical single sorting and consistency check; (4) the level total sorting and consistency check. According to the analytic hierarchy process, the solving flow chart is as show Figure 1.

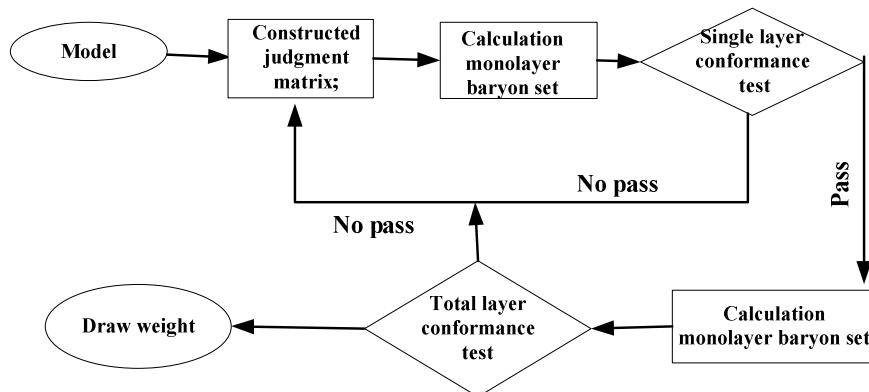


Fig. 1. Analytic Hierarchy Process

3.1 Establish Model of AHP

On the basis of thorough investigation and study, application analytic hierarchy process Analysis principles, layer by layer decomposition of the target, the complex problem into an integral part of the known elements. These elements according to their attributes are into several groups, to formation different levels. The elements of

the same level as a guideline for some elements of the next level play the dominant role, while dominated by a level element. Each floor has a number of elements, the relationship between the layers of elements in a straight line connected [4] [5]. These levels can be roughly divided into three levels: the target layer A, the guidelines layer B , programs layer C; draw this analytic hierarchy process model structure diagram, as shown in Figure 2.

The Target Layer. This level is only one element; the general is the ideal results of the analysis problem.

The Criteria layer. This level include the intermediate links involved to achieve the target, it can be composed by a number of levels, including need to be considered the criteria, sub criteria.

The Program layer. Said for the realization of the goal to choose from a variety of measures, decision-making programs, etc.

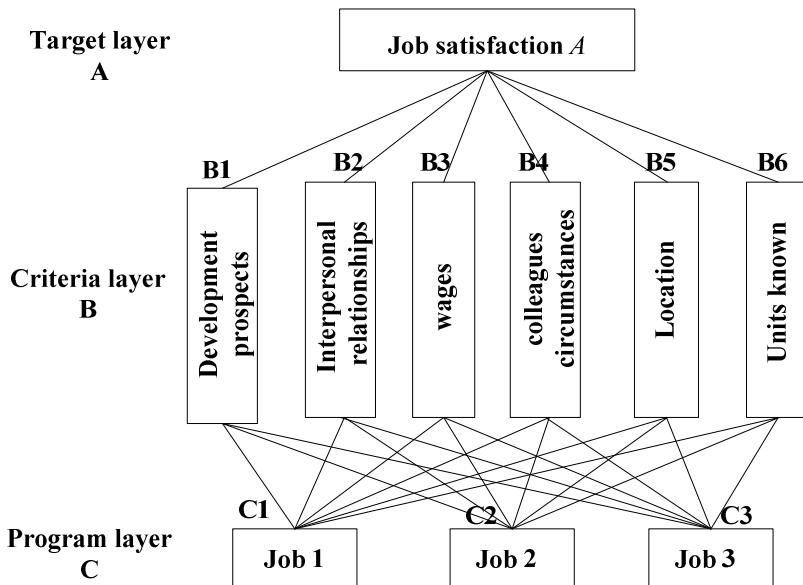


Fig. 2. Analytic hierarchy process model structure diagram

3.2 Construct the Judgment Matrix A

In this paper, using a questionnaire Changchun a university high school student's career investigation, pairwise comparison by the students of the six factors that influence graduates career satisfaction, and after the establishment of hierarchy chart, the system at all levels along with it OK. Then use nine scales TLoSaaty method, in the form of the pairwise proportion of importance to the establishment of pairwise judgment matrix between the rule layer and the various options, the scale is 1 to 9 integers and its reciprocal, called proportional scaling [6]. Which their meanings as show in table 1.

Table 1. The Saaty nine scaling and its mean

Scale	Definition (compare the factors i and j)
1	Factors i and j is equally important
3	Factors i and j is slightly important
5	Factors i and j is strong important
7	Factors i and j is strongly critical
9	Factors i and j is absolutely important
2, 4, 6, 8	Intermediate values of the two adjacent judgments
1 to 9	That factor i and factor j compare the scale value equal to the reciprocal of reciprocal the factor j and factor i compared scale values

After finishing the survey data, to establishing a judgment matrix A, as shown in table 2, the values in the table to determine a student.

Table 2. Structure of the judgment matrix A

A	B1	B2	B3	B4	B5	B6
B1	1	1	1	4	1	1/2
B2	1	1	2	4	1	1/2
B3	1	1/2	1	5	3	1/2
B4	1/4	1/4	1/5	1	1/3	1/3
B5	1	1	1/3	3	1	1
B6	2	2	2	3	3	1

3.3 Level of Single Sorting and Consistency Test

The so-called single-sort is this layer of various factors on the relative importance order of certain factors of the upper. It is expressed by the judgment matrix of eigenvectors. Calculation to determine the matrix A corresponds to the maximum Eigen value λ_{\max} of eigenvectors W . $AW = \lambda_{\max} W$, after by the normalized that the same level of the corresponding factors for the relative importance of some factor of Up one level sort weights, this process is called hierarchical single sort [7]. In this paper, using the data of judgment matrix A available to the index weights accurate estimates in Table 3., Obtain the characteristic vector $W = (0.38, 0.05, 0.21, 0.15, 0.09, 0.17)$ the largest eigenvalue $\lambda_{\max} = 6.2$

Table 3. Index weights accurate estimates Value

A	B1	B2	B3	B4	B5	B6	Weights
B1	4/25	2/23	12/98	3/16	5/28	1/23	0.38
B2	3/25	4/23	13/98	1/4	3/28	5/23	0.05
B3	8/25	5/23	25/98	5/16	7/28	4/23	0.21
B4	3/25	2/23	11/98	1/16	9/28	7/23	0.15
B5	3/25	3/23	19/98	1/8	1/28	1/23	0.09
B6	4/25	7/23	18/98	1/16	3/28	5/23	0.17

In order to ensure the credibility of the hierarchy single sorting, need to test the consistency of judgment matrix, that is, to calculate the random consistency ratio.

$$\text{Consistency index: } CI = \frac{\lambda_{\max} - n}{n - 1}$$

Which n is the order of judgment matrix, when $CI = 0$, the judgment matrix with complete consistency. The larger $\lambda_{\max} - n$, the larger CI , then the judgment matrix consistency backward. Some random consistency values shown in Table 4, the random consistency index use experts summarized the experience value [8].

Table 4. The numerical table of Random Consistency RI

Order	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

$$\text{Consistency ratio: } CR = \frac{CI}{RI} = \frac{\lambda_{\max} - n}{RI \cdot (n - 1)}$$

Only when $CR < 0.1$ the results of the hierarchical single sort are satisfied, otherwise need to adjust the values of judgment matrix elements. In this paper data, $n = 6$, according above formula calculate $CI = 0.04$, View the random consistency RI

values in Table 4. so, $CR = \frac{CI}{RI} = 0.032 < 0.1$, That the establishment of judgment matrix is reasonable to calculate the weight coefficient appropriate judgment matrix by the consistency test[9].

3.4 The Total Level Sort and Consistency Check

Construct all levels of the judgment matrix and its characteristic consistency ratio CR value calculation results and test, only the $CR < 0.1$, the hierarchical single-sort the results considered to be satisfactory, or need to adjust to judge the value of the matrix elements.

3.5 Analysis of Results

It can be seen from the above analysis; six factors of the prospects for development, interpersonal relationships, working conditions, colleagues' situation, geographical location, unit known such as the weights were 0.38, 0.05, 0.21, 0.15, 0.09, and 0.17. Such as in accordance with the right to re-sort, the most important factor is the development prospects followed by working conditions, the unit known, colleagues situation, geographical location, and interpersonal relationships. The most important factor is interpersonal relationships. This shows that college students in the choice of future career, the most important are the prospects for personal development and to meet the individual's material needs, the needs of future development and job value. in the for Students engage process of career concept of education and graduate employment guidance for this situation targeted to capitalize on the trend this time employment rate and to guide the career development of individual students will play an extremely important role [10].

4 Summary

This paper Through the application analytic hierarchy process on the post jobs choice, Undergraduate career choice model is constructed, the establishment of judgment matrix and calculate the respective weights for analysis, not only be able to elect graduates choosing the best job, but also qualitative to quantitative analysis on the jobs , Graduates' Occupation will be more clear.

Acknowledgment. This paper is supported by the “Jilin Provincial Department of Education Scientific and Technological Research Projects [2012] No. 374”.

References

1. Zhang, J.: Students choosing the value orientation of the initial investigation. Lanzhou University (Social Science Edition) 35(5), 146–147 (2007)
2. Peng, G., Li, S., Sheng, M.: Using analytic hierarchy process to determine government performance evaluation index weight. China Soft Science 6, 136–139 (2004)
3. He, J., Huang, J.: Graduate Employment and career behavior research. Jiangxi Education and Research 1, 68–69 (2009)
4. Yu, H., Zhang, J.-F.: From professional values analysis of contemporary college students need characterized. Guangxi University for Nationalities (Natural Science) 5, 68–76 (2003)
5. Jiang, Q., Xie, J., Ye, J.: Mathematical model, vol. 8, pp. 224–244. Higher Education Press, Beijing (2003)
6. State Board of Education Career Guidance Center: Student Career Guidance. Higher Education Press, Beijing (1995)
7. Li, Y.: College graduates job selection based on the analytic hierarchy process. Economic e Update Mid-journals Sum No.145CTrad, 152–153 (September 2010)
8. Wang, L., Xu, S.: Introduction to analytic hierarchy process
9. Lin, H.: Analytic hierarchy process in choosing occupation application. Statistics and Decision (12), 57–59 (2009)
10. Zhang, X.: College Students Occupation Selection based on AHP. Journal of Changsha University (2011)

Study on Information Literacy of College Students with University Library

Hai-bo Yang¹, Yan-yan Zhang², and Ning Yu²

¹ Library of Hebei United University, 063000 Tangshan, Hebei, China

² Jitang College of Hebei United University, 063000 Tangshan, Hebei, China

yhb523@126.com

Abstract. The university library should take full use of resource advantage to meet the service expectation of college students, university libraries should cultivate the students' information awareness, information morals and information abilities from the aspects of the construction of information resource. This paper illustrates the concept of the information literacy, Proposing the significance of training the information literacy and the methods that the University Library offer in improving the information literacy of university students.

Keywords: University library, information Literacy, information education.

1 Introduction

Information literacy has become a social member of the basic viability, but also a learning society to achieve the essential qualities of life-long learning. Information literacy affect the national core competitiveness of the country, decided the fate of the nation's future. Thus, the information society, the overall quality of talent has higher requirements. College students, who want to be in the present information environment based on an invincible position, are bound to make information literacy as personal qualities important part in the training process to cultivate.

2 The Importance of Information Literacy in Training College Students

2.1 To Train Students Information Literacy Has Become an Inevitable Trend of the Information Age

College students are the main use of the information age network groups, and the whole future of national construction and development of the key players in the world, is the backbone of national information capacity, their level of information literacy is directly related to the potential of information society development in China [1], and stamina. However, the information needs of the community is not a

simple transfer of information or user, but with a strong sense of information and able to skillfully use modern information technology, a large number of fragmented information and data summarized and integrated, so there are more structured high information literacy personnel. Information literacy is the information age, people characteristics. College students information education to cultivate the ability of information awareness and information has become an inevitable trend in the information age and the major issues [2].

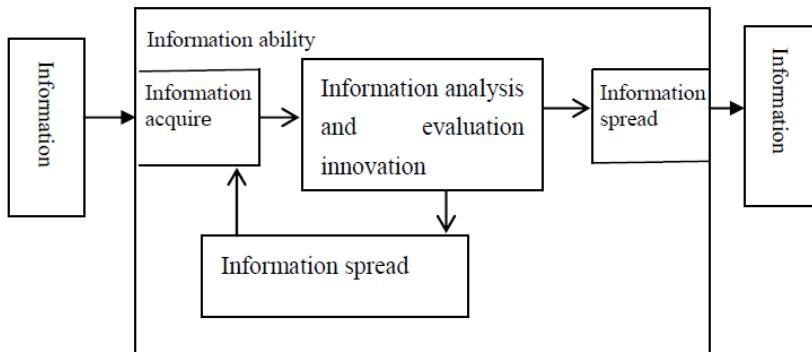


Fig. 1. The methods to obtain information

2.2 To Train Students to Build Information Literacy Is a Fundamental Need for a Learning Society

College Students who are in the golden period of life have active thinking, thinking, studious, sensitive to new things and other physical characteristics, a variety of content on the network to bring them new physical and mental experience, knowledge of rolling waves bring their into a new Information Age, a period of change to give college students a more direct, intense and profound impact. Shoulder to lead students of history, as the master of the history they have to brave social responsibility, constantly enhance their information literacy, learning to survive, learning to create, holding the constant promotion of social development, they are the real masters of the era [3]. Therefore, cultivating information literacy is not only the era of the urgent need to build but also a learning society is a fundamental need.

2.3 Information Literacy Is the Survival and Development of Necessary Skills in a Competitive Society for College Students

College students with information literacy is capable, independent learners, they understand their information needs, positive thinking; their own ability to solve problems is very confident and know what information needs to solve the problem. They can use technology tools to access information, communicate. They work on their own high standards, and produce high-quality products. Students with information literacy is flexible and can adapt to change, able to confidently use all

kinds of information to solve the problem, there is a strong sense of innovation and entrepreneurial spirit. College students with information literacy in a competitive society are essential for the survival and development skills [4].

3 The Concept of Information Literacy

Information literacy was first used by the U.S. Information Industry Association Chairman Paul • cars Chomsky (Paul Zurkowski) in 1974[5], the first time information literacy is defined as "a large amount of information tools and the main source of information to get answers to the problem of technology and skills ", then in turn be interpreted as" when people answer questions in the use of information technology and skills." With the rapid development of information technology, it is the interpretation of information literacy are continuously enriched and expanded that information literacy is the awareness of information ethics guidance and information constraints, the use of computers and information technology, high-speed access, correct evaluation of good use and development of information capabilities, is an extension and expansion of the traditional cultural literacy. They include: a keen sense of information; quick ability to obtain information; correct evaluation and identification information; good at information technology and access to information and working, living, learning the ability to combine; use the information obtained to develop information of capacity. Thus, the complete information literacy should include cultural heritage (knowledge level), information awareness (consciousness), information skills (technical level), information ethics (moral), the four level are complementary and indispensable. Information literacy is a very important capability, with the information literate person is to learn how to learn, to rely on information literacy skills, people are able to achieve expression of ideas, share feelings, expand cooperation, effective thinking, research, decision-making and problem solving, lifelong continuous learning and cooperative learning. Information literacy in today's rapid development of such a technology and information resources is increasingly important as a quality [6].

4 The Role of the University Library in Promoting College Students' Information Literacy

University library is the important place which has a lot of information resources, information retrieval, information networks, and information personnel. University library should be Full use of resources, talent and environmental advantages for students to engage in information literacy education, which should be the future focus of the library.

4.1 The Library's Collection of Rich for Promoting College Students' Information Literacy Provide Material Support

Library has a rich collection of resources is to provide knowledge, which is an important place for the dissemination of information. In recent years, the library was

not only a rich collection of paper documents, but also purchased a large number of electronic documents and a variety of databases [9]. With the popularity of computer networks, the emergence of electronic publications, University Library has been actively building the database and the introduction of electronic carrier literature.

Table 1. The using of library resources

items	The top	Ratio(%)
Business department	Reading room	75.1
Chinese literature	Magazine and newspaper	62.5
Foreign documents	English journal	23.8
Methods channel	Online inquiry	94.7
Information website	SOHU	55.6
Electronic resources	VIP Chinese journal	21.6

Table 2. Commonly used information system and database list

precedence	System database	Ratio (%)
1	VIP Chinese journal	21.7
2	Wangfang data	4.3
3	EI compendex plus	3.6
4	Elsevier Service	2.4
5	VIP foreign database	1.8
6	OCLC FirstSearch	1.8
7	PQDD	0.2

4.2 Qualified Librarians Provide Human Resources for College Students' Information Literacy

Information librarians have accumulated a lot of valuable experience and in various information consulting department has many talents, they are provides human resources and support for college students information literacy education. Implementation of our academic librarians, academic librarians go into the department of the college on a regular basis for undergraduate information organization and promotion, and open platform for subject information services, integrated information service platform for academic subjects related to library resources and services, school-Teaching department of the college as a unit, with research information, training meetings, museum resources, teaching resources, academic journals and more than ten columns, in order to fully support the teaching and research work [8].

4.3 Library Humanities Environment for the Improvement of College Students' Information Literacy Provide the Cultural Atmosphere

The beautiful environment, the strong cultural background and the positive atmosphere of the university library influence in virtually Shared with college students. The

university library should speed up the foundation of information network construction, and build a good learning environment, strengthen open information construction.

5 Some Countermeasures Should Be Adopted by University Library in the Improvement of College Students' Information Literacy Process

5.1 Libraries Should Actively Carry Out Guided Reading [7]

All documentation for the library open to all students, not professional restrictions, students can be based on individual strengths, preferences are free to borrow. Therefore, guided reading plays a very important role, guided reading to carry out a direct impact on the development of library resources, the utilization of library collections and library services work.

First of all, good books are recommended to students, the librarians guide students to read more books, read good books. Promote students in the choice of books, the break of their own professional limitations, broaden horizons and improve their overall quality. Books to the students about the basic method of classification knowledge and library shelving, and guide students to find the books they need. Secondly, the use of campus network means to carry out guided reading. Set up in the campus library website New Books, book reviews and other columns.

5.2 Strengthen Librarians Own Construction, to Improve the Information Service Level of Librarians

Good librarians are an important resource for the library and primary wealth, the university information literacy training and guides. Librarians should have good professional ethics and professionalism, quality service and selfless devotion are the reader plays a subtle role in education. In the information age, librarians of the services relate to various disciplines, professional, must have a basic library and information science knowledge, familiar with the library and collection of business resources, search tools to master the use of a variety of methods, but also modern information technology to enhance training, including basic computer skills, Internet searching skills, and common database structure, retrieval language. In the information environment, librarians should to be keeping pace with The Times, and become a pilot and not the general information service personnel [10].

Libraries should be "Please come to send out" principle, to the library regularly hire senior experts conduct training seminars for the majority of librarians, information librarians to continuously improve the quality and professional level, and select outstanding training librarians to learn and master the most advanced information technology to improve their own quality, in order to better information services for college students to carry out.

5.3 Strengthen Literature Retrieval Course Teaching, Make Full Use of Network Technology to Train College Students' Information Literacy Ability in Practice

Our library' literature retrieval course is the exquisite course, as leaders attach importance to teachers, teaching has been increasing. Document retrieval is generally arranged in sophomore or third grade, students in the library has gradually become familiar with various resources, specialized courses have been opened one after another, the purpose of the practice has gradually increased, to highlight the principles and methods of literature search to guide Students use a variety of foreign language database retrieve relevant information; for higher grades and graduate students to carry out in-depth literature search and use of professional education, focusing on the subject to guide students from the start, begin the subject of thematic analysis, selection and professional database search, evaluation of the results for the search and screening [11].

6 Conclusion

University Library is a means of advanced information services organizations, college students acquire knowledge, improve training, and character cultivation base for educating people is the importance of quality education classroom. With the importance of information literacy education, university libraries will become the university's heart, pouring out through a variety of information, library information literacy education in the university's role will become increasingly important.

References

1. Yin, R.: University library to develop the quality education innovation research. *Medical Informatics Magazine*, 77–79 (2011)
2. Wang, W.: Foreign Information Literacy Research News University Library, 16–17 (2009)
3. Zhang, D.: Discussion on the Strategies for Implementing Information Literacy Education by University Library. *Library and Information Knowledge*, 29–31 (2011)
4. Fan, Z.: Strategies for College Libraries to Improve College Students' Information Quality. *Hebei Normal University of Science and Technology (Social Sciences)*, 23–24 (2009)
5. Wang, Y.: University library information quality education new pattern. *Library and Information Service*, 19–20 (2012)
6. Cheng, J.: Constructing information quality education of auxiliary platform. *The Library Forum*, 147–148 (2008)
7. Zhang, A.: Guangdong university of foreign studies the information quality education survey analysis report. *The Library Forum*, 129 (2010)
8. Li, S.: Discussion on the Functions of University Library in Information Literacy Education for University Students. *Journal of Library*, 38–39 (2011)

9. Yang, G.: Talking about the Information Literacy Education of University Students and the Functions of University Library. *Sci-Tech Information Development and Economy*, 29–30 (2010)
10. Mo, P.: University library under the network environment information quality education college students how to innovation. *The Library Journal of Henan*, 128–130 (2010)
11. Tao, S.: Discussion on the Countermeasure for Strengthening University Students Information Literary Education. *Library Work and Study*, 52–53 (2010)

Comparison of Different Expanded GED-GARCH for Exchange Rate Volatility

Yan Gao^{1,3}, Dongmei Li², and Lihui Zhou¹

¹ College of Science

² Qinggong College, Hebei United University, Tangshan, Hebei, 063009

³ Center for Quantitative Economics of Jilin University, School of Business, Changchun, Jilin Province, China, 130012

Abstract. In this article, we use the data of exchange rate market, to compare different expanded forms of GARCH models based on generalized error distribution (GED), such as EGARCH, TGARCH, PGARCH and CGARCH based on GED. Then we'll find out the best model for describing the volatility of exchange rate market by the evaluation index. Results turn out, for present exchange rate, asymmetric model fits better than symmetrical models, so there exist leverage effect. For all expanded model of GED-GARCH, asymmetric P-GARCH is the best model for estimating the volatility of exchange rate. Asymmetric models is better than symmetric models, which indicates leverage effect have influence for Exchange rate market.

Keywords: GED-GARCH, Exchange Rate, TGARCH, PGARCH, CGARCH.

1 Introduction

Exchange rate plays an important role in national economy and China's economic relations with foreign countries. Volatility of exchange rate is one of significant reason for indeterminacy of macro-economy. It not only influence china's foreign trade, international capital flow and price level, but also affect the Company Performance, economic behaviour and value in microcosmic view. Along with the development of economic globalization and financial integration, China's economy has more relation with other countries. Since July 2nd, 2005, we started to reform the RMB exchange rate system. We began to implement the basis of market supply and demand, with reference to a basket of currencies conditioning, a managed floating exchange rate system. RMB exchange rate would not stare at U.S. dollar any more, more elastic RMB exchange rate mechanism formed. In nowadays, that how to control the exchange rate risk, has important meaning for reality of financial and economical stability.

GARCH models based on Normal distribution can describe the characters of fat-tail with skew, volatility clustering and long memory for financial market well, but as related theories developed, we find GRACH model based on generalized error distribution can describe the problems more precise. For GARCH models, there are many expanded forms. After ARCH and GARCH are proposed, many authors got

a series of expanded form by research: EGARCH, GJR-GARCH, TGARCH, APGARCH, FIGARCH. Chen Zezhong combined GARCH-M and EGARCH analyzed the volatility of stock market. Zhang Siqi used ARCH-M model to study the time series behavior for yield of stock market in china and analyzed the time varying of risk premium.

In this article, we research the forecasting effect of different expanded GED-GARCH model, we find that the expanded form of GED-GARCH models have more precise forecasting ability than expanded form of GARCH model based on normal distribution.

2 GED-GARCH Model and Its Expanded Form

2.1 GED-GARCH

GED-GARCH model is shown below:

$$\begin{cases} \varepsilon_t | I_{t-1} = \sqrt{h_t} \cdot v_t \\ v_t \sim GED(0, 1, r) \\ h_t = \alpha_0 + \beta_1 h_{t-1} + \alpha_1 \varepsilon_{t-1}^2 \end{cases} \quad (1)$$

We denote formula above as

$$y_t = c + \varepsilon_t \sim GED(c, h_t, r) \quad (2)$$

2.2 TGARCH

TGARCH model is proposed by Zakoian (1990), Glosten and Jafanathan, Runkle (1993) separately. Conditional variance is assigned as:

$$h_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \delta I_{t-1} \varepsilon_{t-1}^2 + \beta h_{t-1}^2 \quad (3)$$

In which, I_{t-1} is characteristic variable, when $\varepsilon_{t-1} < 0$, $I_{t-1} = 1$, or, $I_{t-1} = 0$.

In this model, good news ($\varepsilon_{t-1} > 0$) and bad news ($\varepsilon_{t-1} < 0$) has different influence to conditional variance: for good news, it will give a shock of α_1 , bad news will give a shock of $\alpha_1 + \delta$, if $\delta \neq 0$, the information is nonsymmetrical, if $\delta < 0$, we think there exist leverage effect, the main effect of asymmetric is to increase the volatility; if $\delta > 0$, the effect of asymmetric is to decrease the volatility.

2.3 EGARCH

EGARCH model is proposed by Nelson (1991), Conditional variance is assigned as:

$$\ln(h_t^2) = \alpha_0 + \beta \ln(h_{t-1}^2) + \alpha_1 \left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{h_{t-1}} \quad (4)$$

Left of equation is the logarithm of conditional variance, which means the influence of leverage is exponential, not quadratic, so the forecasting value is conditional variance must be non-negative. Existing of leverage effect can be tested by supposing $\gamma < 0$. If $\gamma \neq 0$, effect of shock exists asymmetry.

2.4 PGARCH

In the Power ARCH model, the power parameter of the standard deviation can be estimated rather than imposed, and the optional parameters are added to capture asymmetry of up to order :

$$\sigma_t^\tau = \alpha_0 + \beta \sigma_{t-1}^\tau + \alpha_1 (|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1})^\tau \quad (5)$$

Where $\tau > 0$, $|\gamma| \leq 1$. Note that if $\tau = 2$, $\gamma = 0$, PGARCH model is simply a standard GARCH specification. As in the previous models, the asymmetric effects are present if $\gamma \neq 0$.

2.5 CGARCH

In CGARCH model, mean reversion to a varying level m_t ,

$$\begin{aligned} h_t^2 - m_t &= \alpha_1 (\varepsilon_{t-1}^2 - m_{t-1}) + \beta (h_{t-1}^2 - m_{t-1}) \\ m_t &= \alpha_0 + \rho (m_{t-1} - \alpha_0) + \phi (\varepsilon_{t-1}^2 - h_{t-1}^2) \end{aligned} \quad (6)$$

Here h_t^2 is the volatility, m_t takes the place of α_0 and is the long run volatility of time varying. The first equation describes the transitory component $h_t^2 - m_t$, which converges to zero with powers of $\alpha + \beta$.

The second equation measures the long run component m_t , which converges to ω with powers of ρ . ρ is between 0.99 and 1 so that m_t closes to ω very slowly. We combine the transitory and permanent equations as below form:

$$\begin{aligned} h_t^2 &= (1 - \alpha - \beta)(1 - \rho)\omega + (\alpha + \phi)\varepsilon_{t-1}^2 - (\alpha\rho + (\alpha + \beta)\phi)\varepsilon_{t-2}^2 \\ &\quad + (\beta - \phi)h_{t-1}^2 - (\beta\rho - (\alpha + \beta)\phi)h_{t-2}^2 \end{aligned} \quad (7)$$

which shows that the component model is a GARCH (2, 2) model of nonlinear restricted.

An asymmetric Component GARCH model may be estimated by checking the Include threshold term. This model combines the GARCH model with the asymmetric TARCH model, introducing asymmetric effects in the transitory equation and estimates models:

$$\begin{aligned} m_t &= \alpha_0 + \rho(m_{t-1} - \alpha_0) + \phi(\varepsilon_{t-1}^2 - h_{t-1}^2) \\ h_t^2 - m_t &= \alpha_1(\varepsilon_{t-1}^2 - m_{t-1}) + \delta(\varepsilon_{t-1}^2 - m_{t-1})d_{t-1} + \beta(h_{t-1}^2 - m_{t-1}) \end{aligned} \quad (8)$$

where d is the dummy variable indicating negative shocks. $\gamma > 0$ indicates the presence of transitory leverage effects in the conditional variance.

3 Positive Study and Conclusion

Since July 2nd, 2005, our country began to reform the exchange rate system, so we use daily data after that day, because of American subprime crisis, we consider of getting rid of the influence of the crisis. The exchange rate didn't change much, so we abandon data in this period. Then we use other data describe the volatility of exchange rate. After we take log difference of yield series, we get stable yield series, which is shown in Figure 1. Data comes from Hexun network: <http://forex.hexun.com>.

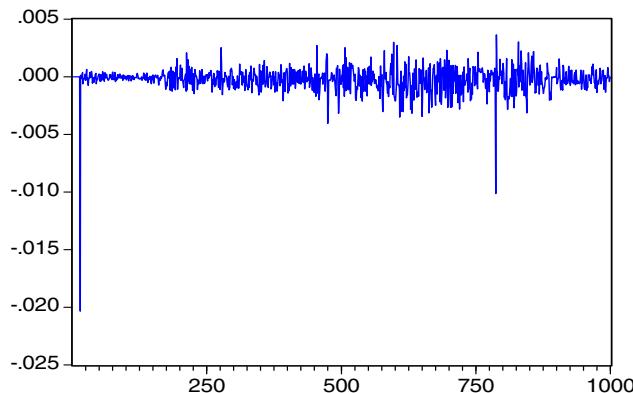


Fig. 1. Yield series of exchange rate

Table 1. The Results of Estimation for different distribution

Coifient	GARCH(1,1)	T- GARCH	E- GARCH	E-GARCH(asy)
α_0	7.52e-09	5.55E-10	-0.158394	-0.106059
α_1	0.133951	0.125390	0.075519	0.130762
β	0.889717	0.972043	0.992586	0.997260
r	0.904116	1.059402	0.970908	0.978070
δ		-0.127342		
γ				0.109540
τ				
ρ				
ϕ				

Table 1. (*continued*)

Coifcient	PGARCH	P-GARCH(asy)	CGARCH	C-GARCH(asy)
α_0	2.03E-05	9.65E-06	3.62E-05	7.90E-07
α_1	0.044087	0.054540	-0.167579	-0.203413
β	0.964915	0.962719	0.273999	0.123884
r	0.950432	0.959237	1.259895	1.531916
δ				0.134362
γ		-0.404182		
τ	0.738253	0.823382		
ρ			0.999031	0.844160
ϕ			0.323756	0.381942

Now we compare the fitted effect for different expanded GED-GARCH, the Akaike information criterion, Schwarz criterion, Sum squared resid and Log likelihood are shown below:

Table 2. Comparison of fitted effect for different expanded GED-GARCH

Coefficient	GARCH(1,1)	T- GARCH	E-GARCH	E- GARCH(asy)
AIC	-11.30958	-11.31055	-11.34999	-11.34966
BIC	-11.28477	-11.28077	-11.32517	-11.31988
SSE	0.001348	0.001350	0.001347	0.001349
Log likely- hood	5580.624	5582.100	5600.543	5601.383
Coefficient	P-GARCH	P- GARCH(asy)	C- GARCH	C- GARCH(asy)
AIC	-11.36120	-11.36138	-11.31255	-11.27004
BIC	-11.33142	-11.32664	-11.27781	-11.23033
SSE	0.001349	0.001174	0.001346	0.001342
Log likely- hood	5607.071	5608.160	5584.088	5564.127

From the effect of fitting, the asymmetric P-GARCH is the best model for estimating the volatility for exchange rate. Then the symmetric P-GARCH is the second better one, which means the PGARCH model is better than other GARCH models, for each model, asymmetric models is better than symmetric models, which indicates leverage effect have influence for Exchange rate market.

Acknowledgments. Supported by: Science research and development guide plan of Tang Shan (No.111102033b); Science research fund of United University (No.z201114), Nature Science Fund of Hebei province, (No.A2010000905; No. A 2012209030).

References

1. Giller, G.L.: A Generalized Error Distribution, Investments Research Note: 20031222/1 (August 16, 2005)
2. Gao, Y., Zhang, C., Zhang, L.: Comparative Analysis of Three GARCH Models Based on MCMC. In: Liu, C., Chang, J., Yang, A. (eds.) ICICA 2011, Part II. CCIS, vol. 244, pp. 494–499. Springer, Heidelberg (2011)
3. Pan, H.: Time Series Analysis. The Press of University of International Business Economics (January 2006)
4. Wang, Z., Hu, Y.: Applied Time Series Analysis. Science Press (May 2007)
5. Cui, C.: An Analysis of the Stock Price Volatility in China Stock Market Based on the GED Distribution. Journal of Ningbo University (Liberal Arts Edition) 22(4), 94–98 (2009)
6. Engel, R.F.: Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 987–100 (July 1982)
7. Tian, Q.-R., Luan, C.-F.: Analysis of Stochastic Volatility Models in Measuring Risk in Chinese Stock Markets. *Science Technology and Engineering* 9(10), 2834–2837 (2009)
8. Bollerslev, T.: Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* 31, 307–327 (1986)
9. Fan, J., Fan, Y., Lv, J.: Aggregation of non parametric estimators for volatility matrix. *Journal of Financial Econometrics* 5(3), 321–357 (2007)
10. Jorion, P.: Value at Risk: The New Benchmark for Managing Financial Risk, 3rd edn. McGraw-Hill (2006)
11. Engel, R.E., Kroner, K.F.: Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11, 122–150 (1995)
12. EViews 5 User's Guide (April 15, 2004), <http://www.eviews.com>

Component Description Model Based on Software Architecture

Zhiyu Chen, Liufu Song, Junqiang Liu, Ming Hu, and Rujia Gao

College of Computer Science and Engineering Changchun University of Technology
huming_ccut@163.com

Abstract. The software modality keeps evolving under the influence of application, platform, and technology development. Internet-ware is a novel software paradigm evolved by the Internet, and the development of Internet-ware has significantly improved the software reuse rate. Architecture-based Component Development is an effective way to improve the development of software. With the development of component technology and the improvement of development methods, it is very essential to describe component clearly for the component user, so people can understand the functions and attributes of component easily. Under the analysis of some component models, we give a description of the component model based on the software architecture in this paper, and this model can provide a method for solving the problem of component composition.

Keywords: internet-ware, component description, component model, component composition.

1 Introduction

After the 1990s, the network has gradually become a new generation of software running environment, compared with the traditional software closed, static, the relative stability of the single operation environment, the network provides an open, dynamic, changeable operating environment with us. And the Internet-ware is rising in this case. The Internet-ware is different from the traditional software; it's an abstraction of the basic form of software systems in the open, dynamic and changeable Internet environment. It is a natural extension of traditional software architecture, but also has unique features that are different from the traditional software form which comes from the focus of a closed environment. Those features' include autonomy, evolution, collaboration, polymorphism and response.

Today, to the software development, the system scale is increasing and becoming larger than before; a system is required to complete a lot of functions and contains more parts. Software reuse and software integration is getting more and more important and extraordinary significance. The traditional software portability and reuse rates are relatively low, if software developed, once delivered, its function has been limited, if you want to modify it, you must pay a great price. In order to increase software productivity and develop effectively application, design and development staff should use the software components as soon as possible, assembled to construct

a new software system. The component technology has accomplished the priority of systematic development from program design to existing software selection, assembling and deployment of conversion.

The reuse of Software is an important direction of the software development; component-based software production has extensive research and application. However, in the traditional manner, software developers in the LAN develop the component library, software component retrieval mechanism, its efficiency and reusability are limited, in order to improve the reusability, we need to improve the traditional way of software development, the internet's open, dynamic, changing operate environment is useful to the component development, it also provides a new development opportunity to the theoretical research of the component. This paper is to describe the component through the analysis of some of the existing component model, and finally we will give a model to describe the component based on the component architecture.

2 Component

Component is a software entity in the software system, this entity has a relatively independent function, and can be clearly identified, the interface of the entity is specified by the contract, to a certain extent, the entity has an obvious dependent relationship with the context; this software entity can be independent deployment and assembled. Each component includes 1 to N interface, and the interface is the key part to the components interaction.

Component is a reusable software module which is facing the software architecture; it can be used to construct other software. Under the function of connectors, some components can be assembled together, and those components can interact together to complete some work that meet the requirements of component users. In order to complete this paper that how to describe the component model, we must give some introductions about the component. The software component is mainly generated from the software requirements of large-scale industrial production. From the point of view of problem solving, Component is facing problems or future problems exists the expectations, it is a basic part of how to solve the problem as widely kind of software system [1].

Component is relatively rapidly developed in recent decades, different component-researchers and organizations have given different definitions to the component. Although the expressions of these definitions are in different forms, they have the same meaning. Component scholars Szyperski gives the definition of component can well reflect the features of the component. The content of the definition as follows: the software component is used for assembly of the unit, it has a standard interface protocol and displays the context dependent, software component can be deployed independently by third party assembly[2].

Component is generally include two parts component interface and component body. Component body is mainly includes the concrete realization of all sorts of function. Component interface provides the basic support for the assembly of components; the component interface is the portal for two component interaction. Due to the loose coupling of the components, component interface plays an important role

in the process of component design. General component interface contains components provide interfaces and component requires external interface provided [3]. Those two interface protocol allow the component to become a relatively independent software entity. About these two types interface, the component provides interface to the external is a necessary, the component users use component to assemble some software is from the component description and the component provide interface. In the component composition to complete a process, connector is using the component provide interface to find some components which fit the needs to complete some work, and through some interface protocols to assemble corresponding components.

3 Component Model

Component model researches the essential characteristics of components and the relationship between components, the component model defines the essential attributes of the components, sets the component interface structures, and the interaction mechanism between components to ontology system structure and components to components. The component model also provides the guiding principle in order to create and realize the component [12]. The typical Internet-ware system is usually composed of various software entities that distributed in the internet, and those software entities are reusable.

The reusability to the reusable object requirements is relatively high. Corresponding, reusable software in the actual construction process we should also consider whether the software object can be fully understood, convenient access, and as well as has versatile functions or not. Therefore in order to make the component-user can easily understand the functions and properties of the component without the need to know how the component was truly completed, give a clear description of the component is necessary. Generally speaking there are three methods to describe the component: models, methods and language. We remain that the component model is the best way to describe the component [4].

Building the model is the focus of the study components, and it is also an important key for the component-user to select the component. Component model is the interaction point of the component research and application is the main basis that the component technology applied to the actual, and plays an important role in the component research field. In order to be able to express the actual situation, a good model should meet the following basic principles:

Completeness component model is an abstract description of software component, is a highly generalizations to component, and there should not have a component which the component model can not describe.

Practicality component model should be easy to understand, facilitate the classification and retrieval.

Expansibility component model can be applied along with the growth of the demand changes, the model is not stabled, and it should have the initiative.

Simplicity simple is an import property that all models should be considered, and it means that the model is easy to understand and grasp. The model is practical, and should not be too complicated.

At present, with the development of research on the component and as well as the depth of the cognitive development, a number of research institutions and organizations give the component model, the more representative, in the country, there are Jade Bird Component model [5] proposed by a research team in Peking University led by Yang Fuqing academicians and Hong Mei professor, Agent model [6] proposed by Lv Jian in Nanjing University etc. They are the guiding model, the abstract level is higher, and the component user can expand the model according to different characteristics of the problem. In the foreign countries, the component model developed relatively quickly, and some of the standard is perfect. Among those models, there are some representative models, such as CORBAR model [7] proposed by Object Management Group (OMG),COM/DCOM/COM+ proposed by Microsoft[8], EJB model[9] proposed by SUN. These models are based on the distributed object technology, affecting the mainstream component technology model.

4 Component Model Architecture

The size and complexity of the software system is still increased, how to effectively develop the large-scale software system is still an issue to the component developers. In recent years, studies have found that based on the reused software is a realistic way [10], architecture-based component assembly is a kind of a reuse-based approach to software development. In this method, component model plays an important role.

In the rapid development of the internet technology, the software is no longer to be thought as a product, but an internet service environment resource. The resource also called component is distributed in the big internet environment, when the server developer confirm the specific services, they can use the component to assemble the service function, In order to correct the expression of the component information, it is required to established an appropriate component model. And we believe that the component can be described by a four topple representation. So we can use the following information to describe the reusable component.

Component can be described by the Component Attribute Description, Component Interface Description, Component Behavior Description and Component Mapping Description. It means that component can be described by: component ::= (CAD, CID, CBD, and CMD).

CAD expresses Component Attribute Description, CID means Component Interface Description, CBD expresses Component Behavior Description, CMD means Component Mapping Description, in the back we will give the specific information that these descriptions represent, and what the description contains.

Component attribute Component attribute is the key to grasp the component, it sums up the component's property, is the starting point for the component users to understand or to use components, it is the component most basic part. It should contain the language how the component developed, the function that the component provides.

Component interface Component has strong loose coupling, the interaction between component and component are not directly, they are through a middleware called Connector, the component interface should include external interface, the interface that needs the external to provide and interface specification. The external interface is provided which must be existed, it contains components external contact information, is the key that interaction between components. However the interface that need the external provide is dependent on the component properties, when the component development did not reach the finish the certain function, it can rely on the interface to change. Interface specification describes the interface behavior, it conducive to the component assembly.

Component behavior Component behavior is determined by the sequence of operation, the message sending and receiving rule. It should include the input information, output information and control information. According to the constitution, component can be divided into atomic component and composite component, atomic component is a simple internal structure, realization is relatively simple. For the composite component the internal structure is more complex.

Component mapping Component mapping description illustrates how the component is achieved, and the CAD is composed of 1 or multiple operating map description. The component is based on the Web reactant as, and the operation positioning mechanism developed the map Web service operations.

Component model depicts the main factors that influent the reusable of component sets the interface mechanism between components to components. Component the designed based on the component model is dependent software entity, in order to construct component system to meet the needs of the user, we should composed those components.

5 Internet Connector

Literature [11-14] points out that the environment of the Internet-ware is dynamic, this means that the component composition workers or the client of the component can not easily assume that the required component is reasonable, that to say the component is loose coupling, and the contact between components is not obvious, the component is an independent and complete individual. This determines the interaction between the components is not directly, the interaction between components is through the interface, so the connector products. Connector is first class entity and used to describe the interaction between components. The internet connector can be generally described by the internal and external protocol. External protocol defines various roles played by the components involved in the component interaction; only those components that meet the role can interact. The internal protocol defines the details how the connector handles the interaction between components, it contains how the components interact, the communication between the interaction, the monitor of the interaction. The connector model is also a black-box model, the connector designer can only give the abstract of the component external interface, and it determines the implementation details of the connector are independent.

6 Summary

This paper presents a component model to describe the component, and the model also gives a concrete realization to meet the situation, the model can summary the network software in some situation, and also provide a way to solve the component heterogeneous problem.

Acknowledgments. Thanks for the support of the Provincial Natural Scientific Fund Project: The Study of the Key Technology of the Web Software Development Platform Oriented to the Service (No. 20101525) and The Project of Ji Lin Provincial Department of Education: The Development Platform based on OSGi(No. 201085).

References

1. Lv, J., Ma, X., et al.: The research and development of internet-ware. *Science in China Ser. E Information Sciences* 36(10), 1037–1080 (2006)
2. Szyperski, C., Gruntz, D., Murer, S.: Component-Software: Beyond Object-Oriented Programming, 2nd edn. Pearson Education Limited. Addison-Wesley/ACM Press, Boston (2002)
3. Jie, Y.: Research on Software Architecture-based Composition of Internet-ware. Peking University Doctor Dissertation (2007)
4. Xu, F., Zhang, A., Huang, H., Wang, Z.: Component Description Model in the Internet Environment (2009)
5. Peking University Department of Computer Science: Jade Bird Software Development Guide, Beijing (1997) (in Chinese)
6. Lv, J., Tao, X., Ma, X., et al.: The development of internet-ware based on agent model. *Science in China Ser. E Information Sciences* 35(12), 1233–1253 (2005)
7. OMG CORBA Specification, <http://www.omg.org/corba/>
8. Microsoft, http://en.wikipedia.org/wiki/Component_Object_Model
9. Oracle, <http://www.oracle.com/technetwork/java/javase/ejb/index.html>
10. Zhang, S., Wang, L., Yang, F.: Architecture Based Software Developing Model World Sci-tech R & D (1999)
11. Ma, H., Zhang, H.: Feature Model-driven Method of Component Composition and Optimization for Internet-ware. *Journal of Chinese Computer Systems* 33(4) (2012)
12. Huang, G., Zhang, L., Zhou, M., Cao, D.: Design and Implementation of Component-Based Software, Beijing (2008) (in Chinese)
13. Wang, J.: Formal Analysis and Construction for Internet-ware-based Software System. Shanghai Jiao Tong University Doctor Dissertation (2009)
14. Ma, H.: Optimization method on component composition for domain feature clustering. *Computer Engineering and Applications* 45(21), 197–200 (2009)

Electronic Transaction Scheme Based on Fast ECC Multiplication

Shouzhi Xu, Bo Xu, Pengfei Guo, and Chenxia Li

College of Computer and Information Technology, China Three Gorges University,
Yichang, 443002, P.R. China
xsz@ctgu.edu.cn

Abstract. Elliptic curve cryptosystem (ECC) becomes the dominating cryptosystem for asymmetric encryption technology applied in e-commerce, information management and office automation systems. ECC is broadly applied to authentication, certification, key distribution and message encryption in recent years. The efficiency of multiplication algorithm is the main impact factors of computation cost for an application system. Combining NAF (non-adjacent form) and variable-length sliding window, an improved variable window mechanism method is presented to scale down the computation complexity of the multiplication of ECC in this paper. Integrated with the improved ECC multiplication algorithm, an electronic transaction scheme is proposed, which includes protocols about opening account, e-cash withdrawal and payment. The proposed scheme ensures security demands, such as non-forgeability of e-cash and avoidance of double spending. Furthermore, both theoretical analysis and test result show that computation efficiency of encryption based on ECC improves well.

Keywords: Electronic Transaction Scheme, Elliptic Curve Cryptosystem, Scalar Multiplication, Variable Sliding-window.

1 Introduction

More and more e-payment scheme apply Elliptic curve cryptosystem (ECC) to provide powerful safety and more efficient computation ability[1], since its quite benefit in stronger safety and computational efficiency than RSA. For example, the security strength of ECC in 160-bit key size is the same with RSA in 1024-bit key size[2]. The registration, authentication, certification, and key distribution in various of security e-payment schemes are broadly computed on devices constrained by low-end processor and memory, such as mobile terminals, smart cards[3].

Unlike traditional public-key cryptosystems which are based on integer factorization problem, such as the RSA algorithm, ECC proposed by Neal Koblitz and Miller in 1985 is based on the difficulty of elliptic curve discrete logarithm problem[4]. The main computation cost of ECC is about its multiplication algorithm. A typical method of NAF (non-adjacent form) is used to reduce computation steps of point multiplication in ECC. A variable window mechanism method of combining

NAF[5] and variable-length sliding window is presented in this paper to scale down the complexity of computations of point multiplication of ECC.

The main contribution of this paper is to propose an e-transaction scheme with the improved NAF multiplication algorithm. The rest of this paper is organized as follows. Firstly, the Weierstrass equation over a prime field $GF(p)$ is described in section 2, its two algorithms including addition and multiplication over a prime field are defined subsequently. Then, an improved algorithm of multiplication NAF(k) based on a sliding window is illustrated. A smart e-transaction scheme with integrating the NAF(k) algorithm is proposed in section 3. Section 4 analyzes the efficiency and the security of the proposed scheme. Finally, the last section gives the conclusion.

2 Improved Multiplication of ECC

2.1 Multiplication of ECC

An ECC system defines an Abelian group that consists of elements over a prime field $GF(p)$ (where p is a large prime), which given by a Weierstrass equation:

$$E : y^2 = x^3 + ax + b \quad (1)$$

Where $p \in (2^k, 2^{k+1})$, $a, b \in GF(p)$, $\Delta = 4a^3 + 27b^2 \neq 0$.

The Abilene group includes a set of pairs (x, y) from function (1) and the point O at infinity. [6] explained more details about operation methods of point addition and point multiplication of $E(F_q)$. Elliptic Curve Discrete Logarithm Problem (ECDLP) is defined as: Given a generator point $G \in E(F_q)$ and any point $R \in E(F_q)$, it is computationally impossible to find an integer k which satisfies $R = kG$.

The efficiency of elliptic curve cryptographic scheme mainly depends on time of scalar multiplication. Scalar multiplication is a computation of the formula $Q = k \cdot P$, which is defined as adding a point P for k times to itself. If binary presentation of k is: $k = (b_i, b_{i-1}, b_{i-2}, \dots, b_1, b_0)_2$. Then the multiplication can be transferred into addition and double multiplication as: $Q = (2(2(\dots(2(b_i \cdot P) + b_{i-1} \cdot P) + \dots + b_1 \cdot P) + b_0 \cdot P)$. Thus, the computation complexity of point multiplication nP reduces to $O(\log(n))$ from $O(n)$.

How to compute kG more efficiently is very important for an encryption system[7]. Since the integer K is always very big in cryptosystem (K is up to 2^{128}), time cost of the multiplication becomes a bottleneck in its application.

2.2 Improved Multiplication Algorithm

Many algorithms of scalar multiplication have been developed such as square-and-multiply algorithm, non-adjacent form (NAF), and sliding window methods[1][5]. The computation time of kP can be divided into point doubling (addition of two same point) and point adding (addition of two different point), whose operation times are dependant on bit length and Hamming weight of the scalar k respectively. But bit

length of the integer k is ascertained, so the execution time of point doubling can not be changed. NAF method is 1/8 faster than square-and-multiply method, the variable-length sliding window method has advantages among sliding window methods.

When combined with variable-length sliding window method, the element of $NAF(k)$ can be divided into two types of windows: zero window and non-zero window. From the left of the element of $NAF(K)$ to the right, window values $\{\pm 1, 0, \pm 1\}$ are selected as non-zero window, whereas window values $\{0, 0\dots 0\}$ as zero window. In this way, it can reduce 1 point addition by combining two non-zero elements. Generally, formula for scalar multiplication can be obtained as follows:

$$kP = 2^{l(E_0)}(2^{l(E_1)}(\dots 2^{l(E_{s-3})}(2^{l(E_{s-2})}E_{s-1}P + E_{s-2}P)\dots) + E_0P \quad (2)$$

$E_i (i = 1, 2, \dots, s-1)$ denotes the i th window, $l(E_i)$ denotes the length of the i th window.

The main steps of the improved multiplication algorithm are shown as Fig. 1:

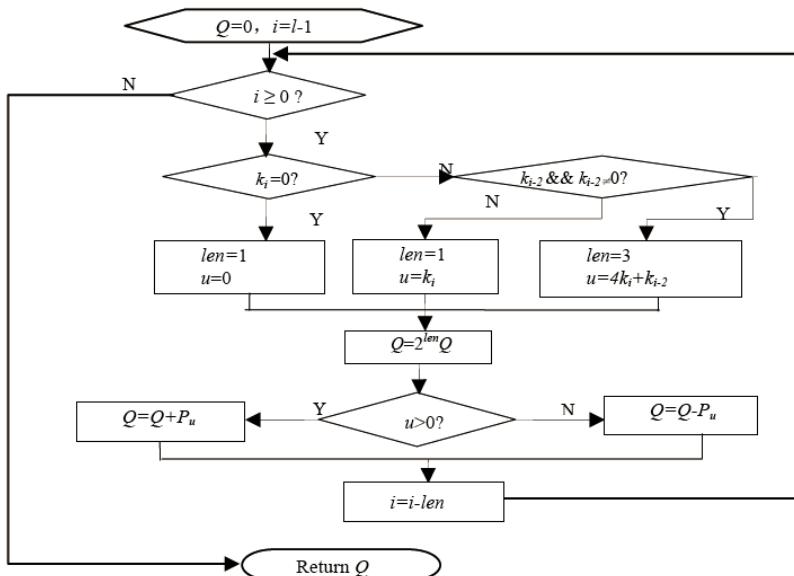


Fig. 1. Point Multiplication

Step 1: Coding k . the integer k is coded into non-adjacent form from $NAF(k) = \sum_{i=0}^{l-1} k_i 2^i$.

Step 2: Calculating kP . Input the parameters(p, a, b, n, P) over $E(F_q)$, then calculates the $Q = kP$ with the above multiplication shown in Fig 1..

The variable-length sliding window definition. Since there are mainly six kinds: $\{1, 0, 1\}, \{1, 0, -1\}, \{-1, 0, 1\}, \{-1, 0, -1\}, \{1\}, \{-1\}$ in non-zero window, window values mainly are $\pm 1, \pm 3, \pm 5$. So the multiplication can be improved by computing $P_1 = 1P$,

$P_3 = 3P$, $P_5 = 5P$ and storing those value (P_i, i) in order to be looked up.

3 ECC Based E-Transaction Scheme

3.1 E-Transaction System Model

The e-transaction system model is shown in Fig.2.

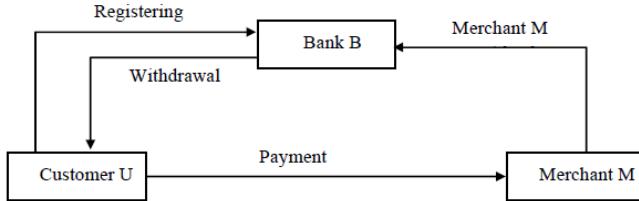


Fig. 2. E-transaction System Model

An e-transaction system is composed of a set of protocols among three participants involved a customer, a merchant and a bank[8] (denoted by U,M and B respectively in this section). In this scheme, the bank B is taken as a trustee[9]. Three fundamental protocols include withdrawal protocol, payment protocol and deposit protocol.

3.2 Registering to Open an Account

B sets up two account databases. The first is for U. The second is for M. B selects $S_1, S_2, S_3 \in_R \mathbb{Z}_q$ secret signature key and computes a corresponding public signature key during set-up $P_1 = S_1 G, P_2 = S_2 G, P_3 = S_3 G$. B sets system bulletin parameters $\{q, a, b, G, n, H(\bullet), Encry, Decry, P_1, P_2, P_3\}$.

U and M open separate accounts with B. B generates a unique secret integer identity for U $ID_U \in_R \mathbb{Z}_q$. B computes and verifies equation for U $pk' = ID_U P_1 + P_2 \neq O$. B computes $\theta_U \equiv ID_U S_1 + S_2 \pmod{q}$ after U accepts, sequentially records $\{name, ID_U, \theta_U, pk', balance\}$ and provides account information to U $Encry(\{name, ID_U, pk', balance\}, S_B P_U)$. B, U and M receive secret signature keys, respectively $\{S_B, P_B\}, \{S_U, P_U\}, \{S_M, P_M\}$.

3.3 The E-Cash Withdrawal Protocol

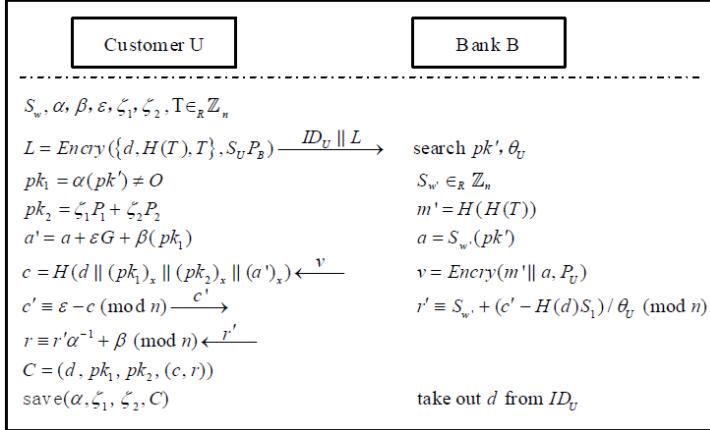
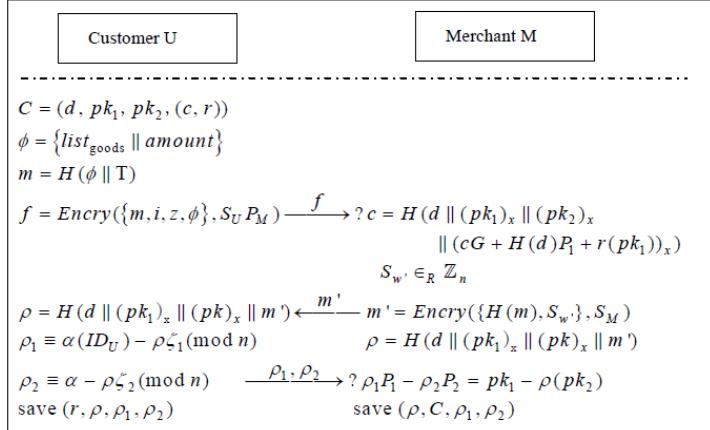
Customer U requests to withdrawal e-cash from bank B in Fig. 3.

U's signature is embedded. B receives a secret message of account balance for U .

U withdraws e-cash $(d, pk_1, pk_2, (c, r))$ with value d ($r \equiv r' \alpha^{-1} + \beta$).U and B sign approval for e-cash signature keys parameters $(\alpha, \zeta_1, \zeta_2)$. U's signature is embedded in e-cash = C, by blinded message c' , and B by r' .

3.4 The Payment Protocol

U purchases e-cash from M. U requests list, amount and quality of goods and sends e-cash to M. The payment protocol is listed in fig. 4.

**Fig. 3.** Withdrawal Protocol**Fig. 4.** Payment Protocol

According to the signatures (ρ_1, ρ_2), the bank B can reveal identity of U if e-cash is spent twice.

3.5 The Deposit Protocol

When the merchant M get an e-cash from the customer U with the payment protocol, M should deposit the e-cash to itself account with a deposit protocol. The merchant M can transmit array of e-cash certificates (ρ, C, ρ_1, ρ_2) and deposit e-cashes in batch from the bank B. If the bank B receive the deposit request, B will check the validity of the e-cashes, if an e-cash has not been spent more than one times and it is in valid life time(expiration date set for each e-cash), the bank B will record relative message and deposit to the merchant B's account. Otherwise, the bank B will refuse to deposit, and the bank will calculate the ID of e-cash master if it is spent more than one time.

4 Security and Efficiency Analysis

4.1 Security Analysis

The primary security of the proposed scheme is as follows:

(1) No one can forge a correct e-cash without secret keys of the bank and a customer, for a legal e-cash needs a set of correct signature, as shown in Fig. 3. Any attacker who wants to get a secret key of other must solve the ECDLP problem.

Anyone can't forge a correct e-cash without secret keys of the bank.

Proof: If an e-cash $C = (d, pk_1, pk_2, (c, r))$ can be used to buy goods following the proposed payment protocol, it must satisfy equality:

$$cG + H(d)P_1 + r(pk_1) = S_{w'}(pk') + \varepsilon G + \beta(pk_1) \quad (3)$$

Since $pk_1 = \alpha(pk')$ and $pk' = ID_U P_1 + P_2$, we have:

$$\begin{aligned} cG - \varepsilon G &= S_{w'}(pk') + \beta\alpha(pk') - r\alpha(pk') - H(d)P_1 \\ &= ((S_{w'} + \beta\alpha - r\alpha)(ID_U S_1 + S_2) - H(d)S_1)G \end{aligned} \quad (4)$$

So $c - \varepsilon = (S_{w'} + \beta\alpha - r\alpha)(ID_U S_1 + S_2) - H(d)S_1$.

But any attacker who wants to get secret keys S_1, S_2 of bank must solve the ECDLP problem described in section 2, so he can't forge such parameters in function (4) to make a correct e-cash C.

(2) If an e-cash is spent more than once, its owner can be discovered by the bank by function $ID_U \equiv (\rho'\rho_1 - \rho\rho'_1) / (\rho'\rho_2 - \rho\rho'_2) \pmod{n}$, in which the set $(\rho', \rho'_1, \rho'_2)$ is subsequent signature generated in payment protocol.

4.2 Efficiency Analysis

Efficiency impact factors of an electronic transaction scheme state mainly from storage requirements and computational cost.

(1) Storage requirements: Storage requirements are critical efficiency factors. No exponential computation scheme is necessary. This allows better efficiency opposed to current providers based on RSA or DSA. ECC's 160 bits surpasses 1024 bits in RSA[10]. [11] proposed a similar scheme based on ECC, its message storage for e-cash is 1464 bits, but our proposed system mandates 960 bits. That is to say, our scheme affords a thirty-percent reduction storage requirement.

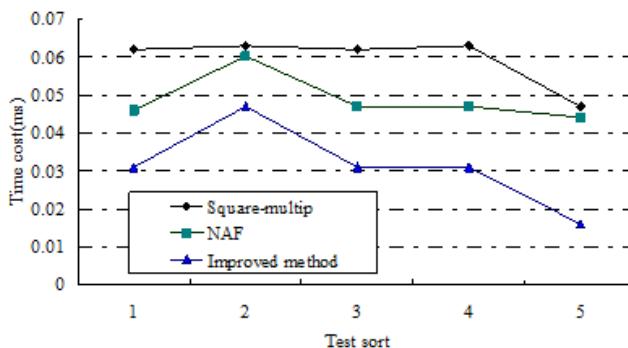
(2) Comparison of computation efficiency: Set m is the bit length of integer K , w denotes the window size, A denotes point addition, D denotes point doubling. 1 A computation includes 1 I (inversion), 1 S (squaring) and 2 M (multiplication); 1 D computation includes 1 I , 2 S and 2 M . On view of computation complexity, 1 I = $30M$, and 1 S = $0.8M$.

Square multiplication is a standard ECC multiplication, our NAFw is improved from NAF, the comparison of these point multiplication algorithms is listed in Table 1.

Table 1. Comparison of Point Multiplication Algorithms

Method	Average Point Addition and Doubling	Computation(m)
Square-multip	$(m / 2)A + mD$	50m
NAF	$(m / 3)A + mD$	44.5m
NAF _w	$1D + (2^{w-2} - 1)A + mA / (w + 1) + mD$	$32.8(2^{w-2} + m / (w + 1)) + 33.6m + 0.8$

NAF method reduces the average number of point addition to $m / 3$, and computation to 44.5m; compared with NAF method, NAF_w method scales down the number of point addition obviously because it combines several non-zero coefficients on the basis of NAF method[6]. In this paper, the presented variable window mechanism combining NAF and variable-length sliding window method with adequate re-computation is more advantageous than NAF_w method, because it can select window size flexibly. So it can improve the execution speed obviously compared with the other previous methods.

**Fig. 5.** Comparison of Computational Efficiency

Signature and verification require 4 times scalar multiplication in the ECC signature scheme totally: 1 key pair generation, 1 signature generation and verification 2, 3 times scalar multiplication. So the method of combining NAF and variable-length sliding window is applied to the implementation of 3 times scalar multiplication based on base point G , and NAF method to the implementation of scalar multiplication of variable point Q .

In order to compare efficiency of several scalar multiplication methods directly, those methods are developed in the same environment with LibTomMath and vc++ 6.0 studio. the result comparison of computational efficiency is shown as Fig. 5.

The result shows that the improved point multiplication need less time, and its efficiency is highest in these existing methods, it is higher 47.6% than the square-multiple method, higher 37.3% than NAF method, and 27.4% than NAF_w method. So both theoretical analysis and test result show that the improved point multiplication algorithm can effectively increase speed.

5 Conclusion

Elliptic curve cryptosystem is a emerging public key cryptosystem which is broadly applied in registration, authentication, certification, or key distribution. To improve its computation efficiency, this paper presented a variable window mechanism method of combining NAF and variable-length sliding window to scale down the complexity of computations of point multiplication of ECC. Both theoretical analysis and test result show that computation efficiency of encryption based on ECC improves well.

The proposal mandates new applications especially mobile electronic payments which support exponential operations ECCDLP. Set protocols and security features afford protection against compromising. Furthermore, the double-spending-resistant system will revoke customer anonymity. The system traces e-coins and customers. Compared with recent research work, our electronic payment schemes have simplified the system structure and act more efficiently.

Acknowledgments. This work was partially supported by National Natural Science Foundation of China (61174177) and Yichang Science and Technology Foundation (A2011-302-13).

References

1. Li, Q., Qiang, Z., Liu, X.: Fast Point Operation Architecture for Elliptic Curve Cryptography. In: IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2008, pp. 184–188 (2008)
2. Ali, R.: Elliptic Curve Cryptography A new way for Encryption. In: International Symposium on Biometric and Security Technologies, pp. 1–5 (2008)
3. Eslami, Z., Talebi, M.: A new untraceable off-line electronic cash system. Electronic Commerce Research and Applications 10(1), 59–66 (2011)
4. Qiu, Q., Xiong, Q.: Research on Elliptic Curve Cryptograph. In: The 8th International Conference on Computer Supported Cooperative Work in Design Proceedings, pp. 698–701 (2004)
5. Wang, B., Zhang, H., Wang, Y.: An Efficient Elliptic Curves Scalar Multiplication for Wireless Network. In: 2007 IFIP International Conference on Network and Parallel Computing Workshops, pp. 131–134 (2007)
6. Fayed, M., Watheq, M., Gebali, F.: A High-Speed, High-Radix, Processor Array Architecture for Real-Time Elliptic Curve Cryptography Over GF(2m). In: 2007 IEEE International Symposium on Signal Processing and Information Technology, pp. 56–61 (2007)
7. Xu, H., Shah, P.G., Sharma, D.: Fast Scalar Multiplication for Elliptic Curve Cryptography in Sensor Networks with Hidden Generator Point. In: 2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 243–249 (2010)
8. Yang, J., Chang, C.: An efficient three-party authenticated key exchange protocol using elliptic curve cryptography for mobile-commerce environments. Journal of Systems and Software 82(9), 1497–1502 (2009)
9. Chen, Y., Chou, J., Sun, H., et al.: A novel electronic cash system with trustee-based anonymity revocation from pairing. Electronic Commerce Research and Applications 10(6), 673–682 (2011)
10. Wang, C., Xuan, H.: A Fair Off-line Electronic Cash Scheme Based on RSA Partially Blind Signature. In: 1st International Symposium on Pervasive Computing and Applications, pp. 508–512 (2006)
11. Popweau, C.: A fair off-line electronic cash system based on elliptic curve discrete logarithm problem. Studies in Informatics and Control 14(4), 291–297 (2005)

Fine Granularity Pipeline Indexing Algorithm for Multi-core Platform

Su-Qing Zhang¹ and Ji-Rui Li²

¹ Information Engineer Department, Henan Vocational and

Technical Institute Associate Professor, Zhengzhou, China

² Information Engineer Department, Henan Vocational and

Technical Institute Lecturer, Zhengzhou, China

Abstract. The scale and growth rate of text collection bring new challenges for index construction. To tackle this problem, an efficient indexing algorithm, the Fine Granularity Pipeline (FGP), is proposed to improve the indexing performance for multi-core platform. Compared to the Simple Pipeline (SP) algorithm, compressing and parsing are divided in the FGP to get more balanced pipeline stages. Evaluations for three collections from Terabyte track in the TREC 2011 over Intel Woodcrest platform showed that the performance improvements of the FGP were over 45% and 17% compared to indexing in Indri 2.4 and the SP with three cores.

Keywords: Indexing Algorithms, Simple Pipeline, Fine Granularity Pipeline.

1 Introduction

We present an efficient indexing algorithm that can be deployed on multi-core systems. It is called the Fine Granularity Pipeline (FGP) algorithm. The FGP algorithm considers the index construction as a pipeline process. It splits the indexing process into several pipeline stages and then runs these stages simultaneously by multi-threading. Since the indexing process is divided and then conquered by execution cores, the FGP algorithm matches multi-core architecture and can achieve good performance.

2 Fine Granularity Pipeline Algorithm (FGP)

2.1 Simple Pipeline Algorithm (SP)

S. Melnik, S. Raghavan, B. Yang and H. Garcia-Molina present a pipelined indexing algorithm for Web search engines. We modified this algorithm so that the new variety can also be applied in traditional information retrieval systems and desktop search environments. We designated this variety as the Simple Pipeline algorithm.

The loading stage loads documents into the memory buffer from the disk or network. During the processing stage, documents are parsed and postings are

extracted from documents. Then postings are processed and accumulated in memory in compressed format, as with the Single-Pass algorithm. During the flushing stage, the compressed postings in the memory are saved on disks as a temporal disk index. These three stages execute repeatedly until all the documents are processed. At last, the temporal indices on the disks are merged into a final inverted index.

The processing stage is CPU-intensive. The loading stage also uses the processor, and the usage depends on the document size. If the document in the text collection is KB scale, then the loading stage may take up many processor cycles. If the document size is MB scale, then the loading stage can benefit from the asynchronous I/O and will use only a few processor cycles. Loading and flushing both use the disk. The flushing stage is activated only when the main memory is used up. In addition, flushing is a burst-write operation and it writes a large amount of data sequentially. Compared with the loading stage, the flushing stage is very fast, so the conflict between loading and flushing is not very serious. If there are two disks in the system, one for the text collection and the other one for the inverted file, the conflict will be minor. It is advantageous to execute these three stages concurrently. Loading, processing and flushing are executed in order so they form a pipeline.

There are some differences between the SP algorithm and the algorithm presented in reference 1. First, the loading stage is extended so that it can read documents from the disk. This is very important for applying the SP algorithm in a desktop environment. Second, in the processing stage of the SP algorithm, postings are compressed, while the algorithm presented in reference 1 stores the original postings in memory. This means the processing stage in the SP algorithm is more like the Single-Pass algorithm, but the processing stage in the algorithm presented in reference 1 is more like the Sort-Based algorithm. Obviously, the former will outperform the latter algorithm, since the Single-Pass algorithm outperforms the Sort-Based algorithm.

2.2 Fine Granularity Pipeline Algorithm (FGP)

The SP algorithm performs better than the Single-Pass algorithm because of the pipelined parallelism. However, the stage division in the SP algorithm is coarse. To solve this, we can divide the indexing process into fine granularity stages to get better performance. This is the key idea of the Fine Granularity Pipeline algorithm.

The stage division in the FGP algorithm is shown in Figure 1. The loading stage is the same as the loading stage in the SP algorithm. The segmentation stage is designed especially for Chinese, Japanese and Korean text collections. During this stage, sentences are segmented into words. This stage is useful only when the text collection contains Chinese, Japanese or Korean text because only sentences in these three languages do not have any delimiters between words. The parsing stage parses documents and tokenizes documents into terms. During the parsing stage, postings are extracted and fed into the following stage - the compression stage. The compression stage processes the posting stream, accumulating lists in the main memory in a compressed format as the Single-Pass algorithm does. When the memory is exhausted, the compressed inverted lists are flushed onto the disk as a temporal disk index in the flushing stage.

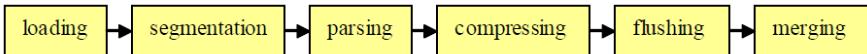


Fig. 1. Stages in the indexing process of FGP

Among these stages, the loading and the flushing stages are IO-intensive. As we mentioned before, the disk conflict between these two stages is not serious. The segmentation stage, the parsing stage and the compression stage are all CPU-intensive. Since the algorithm is employed in a multi-core system, there are multiple execution cores. Each of these three stages can be executed on an individual core for computing. The resource competition between all the stages in the pipeline is not serious, so the pipeline algorithm is suitable for a multi-core system and is expected to have good performance.

The implementation of the FGP algorithm must be threaded to match the multi-core architecture. We can use a dedicated thread for each pipeline stage. For example, a thread dedicated to the loading stage keeps loading documents from the disk into the memory buffer. However, there is another way to implement the Simple Pipeline algorithm and the FGP algorithm. In this alternative, n threads are launched if there are n stages. Each thread is dedicated not to only one stage but to all the stages in the pipeline. Pipelined parallelism is achieved by simultaneous execution of the n threads. However, this alternative has some drawbacks. Because a disk is a sequential device, if one thread is loading documents, the other threads have to wait. If the text collection is stored on a disk, loading documents into the memory by multiple threads will take more time than loading by a single thread because having multiple threads reading simultaneously will harm the spatial locality and sequence of access. Flushing certain amount of data to a disk by several threads is also less efficient than flushing by a single thread. Thus, the dedicated thread approach is more adorable.

For a threaded program, communication between threads is a critical issue for the performance of the program. Main memory buffer is used for thread communication. Generally speaking, the size of the memory buffer is important to the communication efficiency. If the buffer size is too small, it will decrease the communication efficiency. If the buffer size is too big, the buffer will take up too many memory resources. So the goal of the buffer design is to maximum the performance with a minimum buffer size. In the evaluation section we will show that we can get good performance in the cost of a moderate size buffer.

Synchronization is another critical issue for a threaded program. Since the pipeline stages are in sequence, we only have to handle synchronization for adjacent stages. Lock is a common mechanism for thread synchronization. Lock mechanism should be carefully designed because it not only affects the performance of the program, but more important, it affects the correctness of the program.

Careless lock design may cause a very common problem in multi thread program - dead lock. Since adjacent stages communicate by the memory buffer, we can handle thread synchronization in the memory buffer, for example, making the buffer operation thread-safe. When a thread is operating in a buffer, other threads which want to access the buffer at the same time will be blocked. The granularity of the lock is important to performance. If the granularity is too big, other threads will wait a long time to grab a lock. If the granularity is too small, threads will grab and release lock

more frequently and introduce much overhead. For simplicity, in our implementation, the granularity is a document. That means to put a document in the buffer or get a document from the buffer, a thread has to grab and release the lock one time. Of course some other sophisticated lock mechanisms will provide better lock performance, but we can see that even with this simple lock design, the FGP algorithm will outperform the Single-Pass algorithm a lot.

3 Experimental Evaluations

We used three text collections to test our algorithm. These three collections are drawn from the Terabyte track in the TREC 2011. The Terabyte track consists of a collection of Web data crawled from Web sites in the .gov domain during early 2011. This collection ("GOV2") contains a large proportion of the crawlable pages in .gov, including HTML and text, plus the extracted text of PDF, Word, and Postscript files. The GOV2 collection is 426GB in size and contains about 25 million documents. Collection 1, Collection 2 and Collection 3 were disjointed subsets of the GOV2 collection. Since the test collections are English text collection, so we omit the segmentation stage in our indexing process.

The test machine had two Intel Woodcrest 2.66GHz CPUs. Each CPU had four cores, so there were eight cores in the system. However, one of the eight cores had a defect, so we only used the other seven cores in the system. There was 2GB memory in the system and we used 1.5GB memory for constructing the inverted files. The disk was an Ultra320 SCSI disk. And the text collection and the inverted files were placed on the same disk. The operating system running on the test machine was a Linux operating system with kernel 2.4.22.

In the experiments, we measured the performance of the Single-Pass algorithm, the Simple Pipeline algorithm and the FGP algorithm. We implemented these three algorithms and then tested them on a multi-core system.

The Single-Pass program is a modification of *indri* 2.4, which is an efficient indexing and searching engine. The Single-Pass program is composed of three threads. The first thread each time loads a document into the memory, parsing it into postings and compressing the postings. The first thread keeps running until there is no free memory space. Then the first thread is paused and the flushing thread is activated. The flushing thread writes compressed inverted lists in memory onto the disk as a temporal disk index and then it frees the memory. When there is available memory space, the first thread is awakened up and continues to run. When all documents in the text collection are processed, the merging thread merges all temporal disk indices into a final inverted index. The indexing process in the Single-Pass program has a slight difference from the standard Single-Pass algorithm. In the standard Single-Pass algorithm, Golomb codes and Elias codes are used to compress postings. However, use of byte-aligned codes or word-aligned codes can reduce the query evaluation time compared to the Golomb or Elias codes. The overhead of byte-aligned codes is only a modest amount of temporal disk space. Since the word-aligned codes are more complex but have similar performance with byte-aligned codes, for the simplicity, the byte-aligned code is adopted to compress postings instead of the Golomb and Elias codes in the Single-Pass program.

In the Simple Pipeline program, there is a dedicated loading thread. The loading thread loads documents from the disk into the memory buffer. Because documents in the buffer are not parsed, we refer to this buffer as the original document buffer. The second thread fetches documents from the original document buffer and then parses and processes these documents. When the flushing thread is running, the second thread should be paused. However, the loading thread can still run until the original document buffer is full.

In the FGP program, the indexing construction process is divided into four stages. Each stage is implemented by a dedicated thread. The loading thread loads documents into the original document buffer. The parsing thread fetches documents from the original document buffer, parsing these documents and filling the parsed documents into the parsed document buffer. A parsed document is actually a postings array. The compression thread gets parsed documents from the parsed document buffer and compresses them. When the flushing thread is active, the loading thread and the parsing thread can also be active until the original document buffer and the parsed document buffer are full.

The FGP algorithm needs two buffers in the memory. Since the average file size of these three collections is less than 20KB, we test buffer size 512KB, 1MB, 10MB and find that they all have similar performances. Besides, we also generate some bigger documents by aggregating some small documents. The average file size of these bigger documents is 17MB. Then we test buffer size 20M, 50M, 100M and 200M and find that they also have similar performances. So we can achieve good performance in the cost of a little memory for buffering.

Table 1 shows the elapsed time in seconds to construct inverted files with the three algorithms for the three reference collections. The number before the slash is the elapsed time. The percentage after the slash is the performance enhancement compared to the Single-Pass algorithm. For all three collections, when only one core was used, the SP algorithm outperformed the Single-Pass algorithm by more than 24%. The benefit comes from the pipelined execution of the algorithm. However, the indexing time for FGP was a bit longer than for the Single-Pass algorithm. The reason is that there are more threads in the FGP algorithm. The dependence between threads is more complex than the SP algorithm. The overhead introduced by context switching and thread synchronization offset the advantages of the pipeline. Therefore, if there is only one core available in the system, the SP algorithm is more desirable.

Table 1. Elapsed time in seconds to construct inverted files with the three algorithms

Case	Algorithms	1	2	3	4	5
1	Single-Pass	400				
	SP	297/26%	256/36%	257/36%	259/35%	
	FGP	407/-2%	278/30%	207/48%	209/48%	203/49%
2	Single-Pass	1689				
	SP	1265/25%	1091/35%	1076/36%	1082/36%	
	FGP	1721/-2%	1208/28%	923/45%	888/47%	893/47%
3	Single-Pass	3112				
	SP	2362/24%	2016/35%	2014/35%	2042/34%	
	FGP	3253/-4%	2230/28%	1666/46%	1671/46%	1679/46%

When more than one core is used, both the SP algorithm and the FGP algorithm will benefit from these cores. The pipelining time means the time required to process all documents in the text collection, not including the merging time., the processing stage was the stage that had the longest running time. Because the loading stage involves many read operations for files of KB size, the loading stage also uses the processor a lot, so when two cores are used, the loading stage and the processing stage can each occupy their own separate core. Theoretically, the time for a pipeline process is almost the same as the time for the most time-consuming stages in the pipeline. the pipelining time was approximate to the processing time when two cores are used. The processing stage had the longest running time and it hid the times needed by the other stages, so the SP algorithm outperformed the Single-Pass algorithm by 35.3%. Because the flushing stage requires a much shorter time than the loading and the processing stages, an extra core for the flushing stage will not improve the performance much. Therefore, when the number of cores is increased from two to three, the pipelining time does not change much.

In the FGP algorithm, the loading stage also uses the processor a lot, so the loading stage, the parsing stage and the compression stage all need processor resources. When only two cores are used, three stages will compete for the two cores. Context switching and thread synchronization will impair the performance of the algorithm. Thus, when only two cores are used, the FGP algorithm will be less efficient than the SP algorithm. However, when three cores are used, the loading thread, the parsing thread and the compression thread can each be assigned to their own core, so the pipelining time is approximate to the loading time. The performance improvement of the FGP algorithm was 46.5% when three cores were used, which is better than the improvement of the SP algorithm.

Figure 2 shows the relationship between the indexing time and the number of cores used. For the Single-Pass algorithm, the number of cores used did not affect the total indexing time. The SP algorithm outperformed the Single-Pass algorithm significantly when one or two cores were used. However, when the core number continued to increase, it did not improve the performance. For the FGP algorithm, its performance was improved in sequence as the number of cores was increased from one to three. When the number of cores exceeded three, the extra cores did not improve the

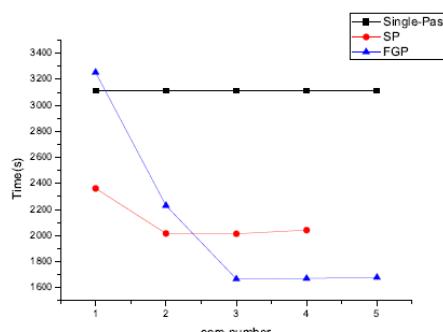


Fig. 2. Scalability of the three algorithms

performance much. Obviously, the FGP algorithm has better scalability than the SP and the Single-Pass algorithm. The reason is that the FGP algorithm has more fine-granularity stages in the pipeline.

However, dividing the indexing process into more stages may not increase the scalability of the FGP algorithm. Therefore, if the parsing stage or the compression stage is divided into more fine-grained stages but the loading stage is left unchanged, the indexing time will not change either. To improve scalability, the running time of the loading stage should be shortened first. Thus, I/O optimization is a key issue for improving the scalability of the FGP algorithm.

4 Conclusions

In multi-core environments, traditional sequential indexing algorithms cannot make use of all the cores in a system. They are also not scalable when the number of cores increases. In this paper, we present an efficient indexing algorithm for multi-core systems: the FGP algorithm. The FGP algorithm adopts the pipeline structure and divides the indexing process into fine granularity pipeline stages. The computing work in the indexing process is divided into several stages so they can run in parallel on execution cores in the system. Stages with many I/O operations can overlap with these computing stages since they use different system resources. The time needed by the short stages is hidden by the longer stages, so the FGP algorithm has significantly improved performance. When two or three cores are used, the performance improvement of the FGP algorithm is 28% or 46% compared to the Single-Pass algorithm.

References

1. Melnik, S., Raghavan, S., Yang, B., Garcia-Molina, H.: Building a distributed full-text index for the web. *ACM Transactions on Information Systems* 19(3), 217–241 (2001)
2. Golomb, S.W.: Run-length Encodings: *IEEE Transactions on Information Theory* IT-12, 399–401
3. Elias, P.: Universal codeword sets and representations of the integers. *IEEE Transactions on Information Theory* IT-21, 194–203
4. Scholer, F., Williams, H.E., Yiannis, J., Zobel, J.: Compression of inverted indexes for fast query evaluation. In: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 222–229 (2002)
5. Anh, V.N., Moffat, A.: Inverted Index Compression Using Word-Aligned Binary Codes. *Information Retrieval* 8(1), 151–166 (2005)
6. Anh, V.N., Moffat, A.: Improved Word-Aligned Binary Compression for Text Indexing. *IEEE Transactions on Knowledge and Data Engineering* 18(6), 857–861 (2006)

Collaborative Filtering Algorithm Based on the Preference List in the Digital Library

Gao Feng¹, Liu Zhiyou¹, and Liu Huixin²

¹ Hebei United University Affiliated Hospital, Tangshan, Hebei, China

² Hebei United University Library, Tangshan, Hebei, China

Abstract. Traditional digital library services are built on the explicit needs of the user. The information needs of the user with specific digital resources associated with key words. This is a passive information retrieval service, only to meet the basic needs of users, not through the user's interest in reading and reading goals to provide targeted services. Through the user log information, we can mine user preferences of different digital resources, to establish a list of user preferences. By the user preference list of association and similarity computation, this paper presents a collaborative filtering algorithm based on user preference list to help readers discover more useful knowledge and information in the mass of digital resources in the digital library system.

Keywords: Preference list, Collaborative filtering, Recommendation algorithms, Digital libraries.

1 Introduction

Digital Library personalized recommendation service is a service user needs, which take the initiative to analyze the user's interests, and then intelligent and efficient for users to find information of interest. Collaborative filtering algorithm is widely used recommendation algorithms, and associated information to recommend a better flexibility for a variety of different application scenarios and field. In nearly a decade, collaborative filtering algorithm has been developed by leaps and bounds, matrix decomposition algorithms [1], the similarity nearest neighbor algorithm [2], the algorithm of Bayesian networks and neural network algorithms and other algorithms. In this paper, digital library resources put forward the concept of a user's preference list [3]. By the user preference list of association and similarity calculation, and proposed a collaborative filtering algorithm based on user preference list.

2 User Preference List

In a digital library, usually the reader through search, recommendation on the site to find the target resource, and one or a group of books, short term or long-term reading, during reading the user to focus more on content and not easy to the score data. In this case, we chose to use the reader's access logs as a data source[4]. At the same time,

consider the use of behavioral information of readers, to give the reader of books of implicit rating information, and then draw the user has read the books ordered preference relations, to expect to give a more accurate recommendation. Therefore, we propose user preference list for collaborative filtering recommendation.

2.1 The Definition of the Data Set

A page of the reading interface to access is part of a possible reading session. The reading session is an abstract access to the reader the book page, a few before and after the time associated with continuous page access will be brought together into a session, can effectively reduce the amount of data, and maintain the mutual independence of two real readers reading behavior. Effective reading sessions only refers to the readers really read the session of reading a book [5].

A reading session, you need to define access the border, in order to determine the number of pages of reading in the reading session. A reading session of the border requires a combination of a specific log data to define, but should note the following:

First of all, more than one reading session can take place simultaneously. This may be the user is comparative reading or complementary reading. The same user on the same book can have more than one reading session.

Secondly, the session time information is meaningful. When a group of pages abstract into a reading session, at the same time we determine the time of the session, which will help to capture the user's recent interest in reading.

Collection of access logs to extract effective reading session is scheduled for:

$$S = \{s_i | 0 \leq i < l\} \quad (1)$$

Where l represents the total number of sessions, each session of a five-tuple (user, book, pages, start, time), initiated by the readers of the session were recorded, read books in the session, read the book pages, the session start time and session time. The user is a collection of readers:

$$U = \{u_i | 0 \leq i < n\} \quad (2)$$

n is the total number of users; the book is a collection of books:

$$B = \{b_i | 0 \leq i < m\} \quad (3)$$

m is the total number of titles; pages is a positive integer; start accurate to the second recording session start time; time is a positive integer, time in seconds, recording session time.

2.2 User Preference List Calculation

From the effective reading session, you can deduce the user to read books these have the potential preference relations. To measure a degree of preference of the readers of a book, need to consider several aspects[6]:

First, the degree of preference should be measured through the reader's act of reading, in general, the book pages read and reading time to be able to decide this book is good or bad.

Second, the readers' interest will change over time, so the reader's preference list with the gradual change in time. The extent of this preference should be flat to reduce as time goes by. This is because the readers' interest generally does not occur dramatic changes.

Third, long-term accumulated preference data should form a stable description, and not reduced to a too low value after a long period of changes.

The definition of readers' preferences for a book as follows:

$$\Psi_{u,b} = \sum_{user(s_i)=u \wedge book(s_i)=b} f(s_i) \quad (4)$$

The greater the value of $\Psi_{u,b}$, the reader is more inclined to love this book. User (s_i) is the session S_i in the user component, book (s_i) is the session S_i in the book component, $f(s_i)$ is defined as follows:

$$\begin{aligned} f(s_i) &= f(pages(s_i), start(s_i), time(s_i)) \\ &= (\alpha_p \cdot pages(s_i)^{\beta_p} + \alpha_t \cdot time(s_i)^{\beta_t}) \cdot g(start(s_i)) \end{aligned} \quad (5)$$

Pages (s_i), start (s_i) and time (s_i) are the pages component, start component and time component in the session s_i . α_p , β_p , α_t and β_t are four parameters used to adjust the reading page number and reading preferences, $g(start(s_i))$ is a time decay factor which value greater than 0 and less than or equal to 1 and g is a function which convert time type parameter into real number, $g(x)$ is defined as follows:

$$g(x) = \max(0.8^{(now-x)/(3 \times 365 \times 24 \times 60)}, 0.8) \quad (6)$$

Here, we hope that the user's preference information three years ago can be attenuated to 0.8, and finally stopped at 0.8. This is because we do not want the user's previous interest will be forgotten with the passage of time. We look forward to the user's interest can be stable description, and recent interest can be highlighted.

Finally, the definition of the reader's preference list is as follows:

$$Pref_u = (b_0, b_1, \dots, b_k), \forall i < j: \Psi_{u,b_i} > \Psi_{u,b_j} \quad (7)$$

3 Collaborative Filtering

Collaborative filtering is a typical use of the collective wisdom of the data mining technology, cannot describe the additional knowledge of the goods or user-assisted recommendation, the basic idea is to find users or items on the similarity of the nearest neighbor, the use of their recommended in the interest on the similarity of the target user has not been contact, but it is widely recognized items [7].

3.1 The Similarity between the Calculations of User

As we adopt an ordered list of models to describe the readers, we need to find an algorithm to measure the similarity of the list. We consider the proportion of common items and the order of common items in two lists. The proportions of common items in two lists reflect the differences in the reading range of different readers. The order of common items reflects the different preference levels of the different readers of the same books.

3.1.1 The Similarity on the Order of Common Books

For the calculation of the similarity on the order of common books, Kendall correlation coefficient is more commonly used[8]. The basic idea is for two ordered lists which the length of the same N elements contained in the same, each with $N \times (N-1)/2$ binary ordered pair, to calculate the simultaneous binary ordered the proportion in the two lists. Kendall correlation coefficient between -1 and 1, 1 is completely similar, -1 is exactly the opposite. The specific formula is as follows:

$$s = \frac{N_c - N_d}{\frac{1}{2} N (N-1)} \quad (8)$$

N_c is the number of tuple of the same order, N_d is the number of tuple of the reverse order. Kendall correlation coefficient applied to the reader model, you can get the similarity of the two readers, the following formula:

$$s_{u,v} = \frac{\sum_{b_i, b_j \in B_u \cap B_v} I((\Psi_{u,b_i} - \Psi_{u,b_j}) \cdot (\Psi_{v,b_i} - \Psi_{v,b_j}))}{|B_u \cap B_v| \cdot (|B_u \cap B_v| - 1)} \quad (9)$$

B_u is a collection of reader u has been reading books, the function I is defined as follows:

$$I(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases} \quad (10)$$

3.1.2 The Similarity on the Proportion of Common Books

Suppose we need to calculate the similarity of reader u and reader v. Reader u is the main. Reader v is the object. For the calculation of the similarity on the proportion of common books, we need to consider two factors [9]:

First, the greater the proportion of common books in the collection of the reader v has been reading books, the closer the reading range of the reader v is to the reading interest of the reader u within the scope of the common books. This will give the recommended results of the reader u greater help. Conversely, the smaller the proportion of common books, the farther the reading ranges of the reader v is to the reading interest of the reader u. The help from the reading range of the reader v is reduced accordingly.

Second, the greater the proportion of common books in the collection of the reader u has been reading books, the closer the reading interest of the reader u is to a subset of the reading interest of the reader v . Predict a subset of the complete works to be more accurate. Conversely, the smaller the proportion of common books, the farther the reading interest of the reader u is to the reading interest of the reader v . The help from the reader v is reduced accordingly.

Calculation of the factors on the similarity of the two readers in the following way:

$$\text{Sim}_{u,v}^{\text{cb}} = \frac{|B_u \cap B_v|}{|B_u|} \cdot \frac{|B_u \cap B_v|}{|B_v|} = \frac{|B_u \cap B_v|^2}{|B_u| \cdot |B_v|} \quad (11)$$

3.1.3 Comprehensive Similarity

Two factors mentioned above, we finally define the similarity of the two readers:

$$\begin{aligned} \text{Sim}_{u,v} &= s_{u,v} \cdot \text{Sim}_{u,v}^{\text{cb}} \\ &= \left(\frac{\sum_{b_i, b_j \in B_u \cap B_v} I((\Psi_{u,b_i} - \Psi_{u,b_j}) \cdot (\Psi_{v,b_i} - \Psi_{v,b_j}))}{|B_u \cap B_v| \cdot (|B_u \cap B_v| - 1)} \right) \frac{|B_u \cap B_v|^2}{|B_u| \cdot |B_v|} \end{aligned} \quad (12)$$

Obviously, the value of $\text{Sim}_{u,v}$ is in the range between -1 and 1.

3.2 Predict Users' Preferences of Unknown Books

When you get a list of all user preferences, as well as the similarity between any two users, you need to use these two aspects of the information to predict the users on the list of preferences of those who read books to pass the list of predicted to recommend books.

3.2.1 Predict the Partial Order Relation between Any Two Books

Directly through a number of weighted ordered lists to predict an ordered list is more difficult. Therefore, the re-combination of the list is used to predict the final ordered list. The first step is the decomposition of an ordered list. The second step is to forecast contained in these lists books, preference relations of any two books, the strength of this preference relation.

For any user's preference list, Pref_u , is decomposed into a partial order set:

$$P_u = \{(B_i, B_j) : B_i \in \text{Pref}_u \wedge B_j \in \text{Pref}_u \wedge B_i \neq B_j\} \quad (13)$$

Books to define the need to predict the range of:

$$\hat{B}_u = \bigcup_{\text{Pref}_u \cap \text{pref}_v \neq j} \text{Pref}_v \quad (14)$$

And the need to predict the partial ordering of collection:

$$\hat{P}_u = \left\{ (B_i, B_j) : B_i \in \hat{B}_u \wedge B_j \in \hat{B}_u \wedge B_i \neq B_j \right\} \quad (15)$$

In accordance with the classic Item-based collaborative filtering algorithm [10], forecasting is a partial order relationship between the collection of any two books:

$$\hat{i}_u(b_i, b_j) = \frac{\sum_{v \in Neib_u} \sum_{b_i, b_j \in B_u \cap B_v \wedge b_i \neq b_j} i_{b_i, b_j}^v \cdot Sim_{u,v}}{\sum_{v \in Neib_u} |Sim_{u,v}|} \quad (16)$$

$Neib_u$ is the reader u "neighbors". The reader u read at least one reader of this book.

3.2.2 Merge Partial Order Relation Set

When we get a set of predictable set of partial order relations, it needs to be combined into a complete list. Predicted the partial ordering of set does not necessarily deduce a consistent sequence table, therefore, we should determine the goal of a combined to determine the combined list matches the extent of the collection of the original partial order relation. Define an evaluation function as follows:

$$V^{\hat{i}_u} = \sum_{b_i, b_j : P(b_i) < P(b_j)} \hat{i}_u(b_i, b_j) \quad (17)$$

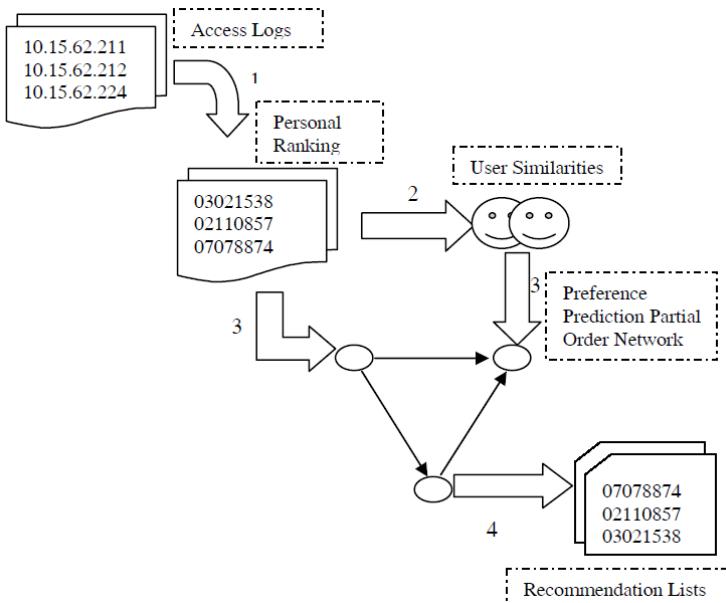


Fig. 1. Collaborative filtering algorithm based on the preference list

$P(b_i)$ is b_i precedence in the combined sequence table, $P(b_i) < P(b_j)$ means b_i precedence before b_j , in the combined sequence table. The larger the value of $V^{\hat{i}_u}$, the combined sequence table is more in line to predict the partial ordering of a collection.

3.3 Generation Were Recommended

Get a prediction order preference lists, you can easily get the recommended results. The methods were: to get rid of the books which the readers have read from totally ordered preference list, to keep other books of the same order, the first K books were recommended. Finally, the process of the entire algorithm is shown in Figure 1:

4 Conclusion

Collaborative filtering algorithm based on the preference list in a digital library is described in detail. This paper presents the concept of the preference list, the reader's access logs to generate the reader's preference list, and then introduced the similarity calculation between readers based on the preference list, and how to use the similarity between the readers to predict the reader preferences between any two books in the candidate list of books. Finally, the predicted preference relation of partial ordered sets combined into an ordered list of preferences to get the recommended results.

References

1. Yu, L., Liu, L., Li, X.: A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce Original Research Article. *Expert Systems with Applications* 1(28), 67–77 (2005)
2. Takacs, G., Pilaszy, I., Nemeth, B., Tikk, D.: Scalable Collaborative Filtering Approaches for Large Recommender System. *Journal of Machine Learning Research* 1(10), 623–656 (2008)
3. Bo, X., Peng, H., Fan, Y.: A distributed collaborative-filtering neighbor-locating algorithm. *Information Sciences* 6(177), 1349–1363 (2007)
4. Kim, H.-N., Ji, A.-T., Ha, I.: Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation. *Electronic Commerce Research and Applications* 1(9), 73–83 (2010)
5. Renda, M.E., Straccia, U.: A personalized collaborative Digital Library environment: a model and an application. *Information Processing & Management* 1(41), 5–21 (2005)
6. Porcel, C., Moreno, J.M., Herrera-Viedma, E.: A multi-disciplinar recommender system to advice research resources in University Digital Libraries. *Expert Systems with Applications* 10(36), 12520–12528 (2009)
7. Gao, F., Xing, C., Du, X., Wang, S.: Personalized Service System Based on Hybrid Filtering for Digital Library. *Tsinghua Science & Technology* 1(12), 1–8 (2007)

8. Forsati, R., Meybodi, M.R.: Effective page recommendation algorithms based on distributed learning automata and weighted association rules. *Expert Systems with Applications* 2(37), 1316–1330 (2010)
9. Krestel, R., Fankhauser, P.: Personalized topic-based tag recommendation. *Neurocomputing* 1(76), 61–70 (2012)
10. Li, M., Liu, L., Li, C.-B.: An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems. *Expert Systems with Applications* 7(38), 8586–8596 (2011)

Customer Threshold Strategies in Observable Queues with Partial Information of Service Time

Wei Sun and Shiyong Li

School of Economics and Management, Yanshan University,
Qinhuangdao 066004, China
wsun@ysu.edu.cn, shiyongli@yahoo.cn

Abstract. This paper concerns customers' decision process of joining or balking at their arrival instants in an observable queue. Given partial information that different number of moments of the service time is available to the arriving customers, they need to estimate its distribution by using only the information available and no other information based on the maximum entropy principle to decide whether join or balk. Respectively for the systems with risk-neutral and risk-averse customers, we numerically find that in general the equilibrium thresholds are no less than the socially optimal one which is also no less than the profit-maximizing one.

Keywords: Observable queue, Partial information, Equilibrium, Social optimization, Profit maximization.

1 Introduction

Recently, much excellent literature on the relationship between the delay information and customers' strategic behavior came forth. Guo and Zipkin (2009) [1] considered two types of vague information about delays. In the first one with partition information, customers learn a rough range of the current queue length, whereas in the second one with phase information, customers learn the total number of phases remaining in the system. Then Armony et al.(2009) [2] investigated the performance impact of making delay announcements to arriving customers in a many-server queue setting with customer abandonment. Furthermore, Sun et al.(2010) [3] assumed both the fully and almost observable cases and derived customers' equilibrium threshold strategies and analyze the stationary behavior of the systems.

All the above works study information on queues. However, we study the information on service times. Considering very often the server doesn't convey the full information about the service time he has acquired to customers deliberately or because of objective factors, to make decision of joining or balking at their arrival instants, the customers needs to estimate the probability distribution of service time by using only the information available and no other information. Guo et al.(2011) [4] first studied the unobservable queue with partial information on service time, while this paper discusses customers' threshold strategies in the observable queue.

For the estimation method, here we assume the customers all conform with the maximum entropy principle. To acquire more, the beginner can consult the literature given by Kapur and Kesavan (1992) [5] where an excellent and detailed introduction about the maximum entropy principle is presented. Here the objective of the maximum entropy principle is to choose, under given information, the maximum entropy distribution (MED) in all possible compatible distributions. Based on the maximum entropy principle and by use of Lagrange undetermined multiplicator method, we may get the MED consistent with the given information. Therefore, the MED of the service time gives customers the most unbiased and most objective distribution consistent with the partial information.

We first derive the equilibrium thresholds under the partial information that different number of service time moments for both the systems with risk-neutral and risk-averse customers. Especially given the actual service time distribution is exponential, we derive the socially optimal and profit-maximizing thresholds for two risk-averse cases and then numerically compare all kinds of thresholds, and find that it is kept in general for the risk-neutral and risk-averse customers that the equilibrium threshold(s) is (are) no less than the socially optimal which is also no less than the profit-maximizing. Then we summarize that the equilibrium threshold under the partial information of mean and variance is greater than that under the partial information of mean.

2 Equilibrium

First assume service reward R and unit waiting cost c if any customer joins, then a Poisson arrival process with parameter λ to a single server system and a tagged customer can observe the queue length $n - 1$ at his arrival instant, whereas he is not told about the exact distribution function F_S of the service time S except some partial information—some first moments of the service time. Denote F as the distribution function of the customer's sojourn time $T = S_1^- + S_2 + \dots + S_n$, where S_1^- is the residual service time of the customer who is just receiving service when the tagged customer arrives. Specially given exponentially service time, $\{S_1^-, S_k, 2 \leq k \leq n\}$ are all i.i.d. random variables which have the same distribution with S , that is, F is the n -fold convolution of the service time distribution function F_S .

2.1 Risk-Neutral Customers

First consider a system with risk-neutral customers, in which any customer's residual utility of joining U_{join} , maybe positive or negative, can be expressed as

$$U_{join} = \int_0^\infty (R - ct)dF(t) = R - cE[T] = R - c(n - 1)\bar{s}_1 - c\frac{\bar{s}_2}{2\bar{s}_1}, \quad (1)$$

where $E[S] = \bar{s}_1$ and $E[S^2] = \bar{s}_2$. As for $E[T]$, because $E[S_1^-] = \bar{s}_2/(2\bar{s}_1)$, we have $E[T] = (n - 1)\bar{s}_1 + \bar{s}_2/(2\bar{s}_1)$. Obviously should assume the customer's

residual utility of balking $U_{balk} = 0$. If $U_{join} > U_{balk} = 0$, then the arriving customer will join the queue. Otherwise, the arriving customer will balk.

If arriving customers only know the mean \bar{s}_1 of the service time, then the following lemma shows the form of the MED based on the partial information (see [5]).

Lemma 1. *If the service time is positive with its mean \bar{s}_1 , then the MED of the service time is exponential with parameter $1/\bar{s}_1$.*

Then the equilibrium threshold strategy is given in the following proposition.

Proposition 1. *If risk-neutral customers are informed of the mean \bar{s}_1 of the service time in an observable queue, then the equilibrium threshold $n_{eM} = \left\lfloor \frac{R}{c\bar{s}_1} \right\rfloor$. If $0 \leq n < n_{eM}$, they will join. Otherwise, they will balk.*

If the partial information is \bar{s}_1 and \bar{s}_2 (or the variance $\sigma_s^2 = \bar{s}_2 - \bar{s}_1^2$), then the equilibrium threshold is obviously obtained by solving $U_{join} = 0$ and it is independent of the actual service time distribution.

2.2 Risk-Averse Customers

Then we consider a system with risk-averse customers, in this case any customer's residual utility of joining U_{join} can be expressed explicitly as

$$U_{join} = \int_0^\infty (R - ct^{m+1})dF(t), \quad \forall m \in N^+, \quad (2)$$

where m indicates the sensitive degree of the risk-averse customers to the potential joining risk. Obviously, customers are classified by their risk sensitivity that increases along with the value of m .

We assume the precondition that the actual distribution of the service time exists any order raw moment. If the arriving customers know the first l raw moments $\{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_l\}$, $1 \leq l \leq m + 1$ of the service time, where $E[S^k] = \bar{s}_k$, $1 \leq k \leq l$, then the following lemma shows the MED of their sojourn time based on the partial information (see [5]).

Lemma 2. *If risk-averse customers are informed of the first l raw moments $\{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_l\}$ ($1 \leq l \leq m + 1$) of the service time in an observable queue, then the density function of MED of their sojourn time is*

$$f(t) = e^{-\lambda_0 - \lambda_1 t - \lambda_2 t^2 - \dots - \lambda_l t^l}, \quad (3)$$

and $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_l$ are the solutions of the following equations:

$$\begin{cases} e^{\lambda_0} = \int_0^\infty e^{-\lambda_1 t - \lambda_2 t^2 - \dots - \lambda_l t^l} dt \\ E[T^k]e^{\lambda_0} = \int_0^\infty t^k e^{-\lambda_1 t - \lambda_2 t^2 - \dots - \lambda_l t^l} dt, \quad k = 1, 2, \dots, l \end{cases} \quad (4)$$

where (we denote $\bar{s}_0 = 1$)

$$\begin{aligned} E[T^k] &= E[(S_1^- + S_2 + \cdots + S_n)^k] \\ &= \sum_{\sum_{i=1}^n k_i=k} C_k^{k_1} C_{k-k_1}^{k_2} \cdots C_{k-\sum_{j=1}^{n-1} k_j}^{k_n} E[S_1^{-k_1} S_2^{k_2} \cdots S_n^{k_n}] \\ &= \sum_{\sum_{i=1}^n k_i=k} \frac{k!}{k_1! k_2! \cdots k_n!} \prod_{j=2}^n \bar{s}_{k_j} E[S_1^{-k_1}] \end{aligned} \quad (5)$$

and $E[S_1^{-k_1}]$ can be solved by differentiating the Laplace-Stieltjes transform (LST) of the residual service time $S^{-*}(s) = (1 - S^*(s)) / (\bar{s}_1 s)$.

Proposition 2. *If risk-averse customers are informed of the first l raw moments $\{\bar{s}_1, \bar{s}_2, \dots, \bar{s}_l\}$ ($1 \leq l \leq m+1$) of the service time in an observable queue, then the equilibrium threshold n_{eM} is the floor function of the solution of*

$$R - c \int_0^\infty t^{m+1} e^{-\lambda_0 - \lambda_1 t - \lambda_2 t^2 - \cdots - \lambda_l t^l} dt = 0, \quad (6)$$

where $\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_l$ are the solutions of Eqs.(4) and (5).

3 Social Optimization

Now we respectively consider the socially optimal thresholds of risk-neutral and risk-averse customers if the actual service time distribution is exponential with parameter μ .

We first denote the expected social welfare per time unit by SW and socially optimal threshold by n^* . Given a maximum queue length n , the system is actually an M/M/1/n queue and the stationary state probabilities $\{p_i, i = 1, 2, \dots, n\}$ are $p_i = \rho^i p_0, \rho = \frac{\lambda}{\mu}, i = 1, 2, \dots, n$, where $p_0 = (1 - \rho) / (1 - \rho^{n+1})$. So the probability of observing n customers in the system, that is, the system loss probability, is

$$p_n = \rho^n p_0 = \frac{(1 - \rho)\rho^n}{1 - \rho^{n+1}}.$$

3.1 Risk-Neutral Customers

Given the partial information is $\bar{s}_1 = 1/\mu$, according to Proposition 1, we have $n_{eM} = \left\lfloor \frac{R\mu}{c} \right\rfloor$. The social welfare per time unit, denoted by SW , is

$$\begin{aligned} SW &= \lambda(1 - p_n) \left(R - \frac{c}{\mu(1 - p_n)} \sum_{i=1}^n i p_{i-1} \right) \\ &= \frac{\lambda(1 - \rho^n)}{1 - \rho^{n+1}} \left(R - c \left(\frac{1}{\mu(1 - \rho)} - \frac{n\rho^n}{\mu(1 - \rho^n)} \right) \right). \end{aligned} \quad (7)$$

So the socially optimal threshold n^* is the floor function of the maximizer of Eq.(7).

3.2 Risk-Averse Customers

Given risk-averse customers, the social welfare per time unit SW is

$$SW = \lambda(1 - p_n) \left(R - c \int_0^\infty t^{m+1} dF(t) \right) = \frac{\lambda(1 - \rho^n)}{1 - \rho^{n+1}} (R - cE[T^{m+1}]), \quad (8)$$

where

$$E[T^{m+1}] = \frac{1}{\mu^{m+1}} \sum_{\sum_{i=1}^n k_i = m+1} \sum_{\sum_{i=1}^n k_i j_{k_i} = m+1} \prod_{i=1}^{m+1} (k_i!)^{j_{k_i}}.$$

Special Case: $m = 1$

- **Mean.** Based on Lemma 1, given the partial information is the mean $\bar{s}_1 = 1/\mu$ of the service time, then the MED is exponential that just coincides with the actual distribution, that is, the customers' estimation is accurate. So the MED of the sojourn time is Erlang distribution with parameter n and $1/\bar{s}_1$, then we get $U_{join} = R - cn(n + 1)\bar{s}_1^2$. Solving $U_{join} = 0$, we get the equilibrium threshold as

$$n_{eM} = \left\lceil -\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{R\mu^2}{c}} \right\rceil. \quad (9)$$

On the other hand, when $m = 1$ the expected social welfare per time unit is

$$\begin{aligned} SW &= \lambda(1 - p_n) \left(R - \frac{c}{\mu^2(1 - p_n)} \sum_{i=0}^{n-1} (i+1)(i+2)p_i \right) \\ &= \frac{\lambda(1 - \rho^n)}{1 - \rho^{n+1}} \left(R - c \left(\frac{1}{\mu^2(1 - \rho^n)} \left(\frac{2\rho^2(1 - \rho^{n-2})}{(1 - \rho)^2} - 3n\rho^n \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{\rho + \rho^{n+1}(n-1)^2 - \rho^n(n^2-2)}{1 - \rho} \right) + \frac{2 + \rho}{\mu^2(1 - \rho)} \right) \right), \end{aligned} \quad (10)$$

and the socially optimal threshold n^* is the floor function of the maximizer of Eq.(10).

- **Mean and Variance.** Given the partial information is both the mean $\bar{s}_1 = 1/\mu$ and variance $\sigma_s^2 = 1/\mu^2$ of the service time, then it is enough for the customers to make decisions so that the equilibrium threshold n_{eMV} is equal to that given in Eq.(9).

4 Profit Maximization

Besides the equilibrium and socially optimal thresholds, we also consider the profit-maximizing threshold denoted by n_m , i.e. the desired threshold chosen by the server. For the risk-neutral case, Hassin and Haviv (2003) [6] presented

that the profit-maximizing threshold n_m is the floor function of the solution of equation

$$\nu + \frac{(1 - \rho^{\nu-1})(1 - \rho^{\nu+1})}{\rho^{\nu-1}(1 - \rho)^2} - \frac{R\mu}{c} = 0.$$

Then we try to derive the profit-maximizing thresholds for two risk-averse cases and consider $m = 1$. Conformed with the threshold n_m , the price set by the server, denoted by p_m , is

$$p_m = R - cE[T^2] = R - \frac{cn_m(n_m + 1)}{\mu}.$$

So the server's profit, denoted by P_{n_m} , is

$$\begin{aligned} P_{n_m} &= \lambda(1 - p_{n_m})p_m = \frac{\lambda(1 - \rho_m^n)}{1 - \rho^{n_m+1}} \left(R - \frac{cn_m(n_m + 1)}{\mu} \right) \\ &= \frac{\lambda R(1 - \rho_m^n)}{1 - \rho^{n_m+1}} \frac{\nu_e(\nu_e + 1) - n(n + 1)}{\nu_e(\nu_e + 1)}, \end{aligned} \quad (11)$$

where $\nu_e(\nu_e + 1) = R\mu^2/c$. Because a profit-maximizing threshold satisfies the following two conditions: $P_{n_m} > P_{n_m-1}$ and $P_{n_m} \leq P_{n_m+1}$ ^[7]. Substituting Eq.(12) into the first condition, then it amounts to $\nu_e(\nu_e + 1) > f_1(n_m)$, where

$$f_1(n_m) = n_m(n_m + 1) + \frac{2n_m(1 - \rho^{n_m-1})(1 - \rho^{n_m+1})}{\rho^{n_m-1}(1 - \rho)^2}.$$

Substituting $n_m + 1$ for n_m and reversing the direction of the inequality, the second condition becomes $\nu_e(\nu_e + 1) \leq f_1(n_m + 1)$, where

$$f_1(n_m + 1) = (n_m + 1)(n_m + 2) + \frac{2(n_m + 1)(1 - \rho^{n_m})(1 - \rho^{n_m+2})}{\rho^{n_m}(1 - \rho)^2}.$$

Then the two conditions can be summarized to $f_1(n_m) < \nu_e(\nu_e + 1) \leq f_1(n_m + 1)$. Define a function

$$f_1(\nu) = \nu(\nu + 1) + \frac{2\nu(1 - \rho^{\nu-1})(1 - \rho^{\nu+1})}{\rho^{\nu-1}(1 - \rho)^2} - \nu_e(\nu_e + 1), \quad (12)$$

which monotonously increases from 0 to ∞ with ν . So there exists a unique solution ν_m^1 to the equation $f_1(\nu) = 0$ and then the profit-maximizing threshold $n_m = \lfloor \nu_m^1 \rfloor$ when $m = 1$.

Figures 1–2 numerically make comparisons of the equilibrium threshold(s) n_e and socially optimal threshold n^* as well as profit-maximizing threshold n_m when $m = 0, 1$ respectively. They jointly show that the three types of thresholds all decrease along with the increase of m . Moreover, Figure 1 verifies the relationship pointed out by Naor (1969) [7] that $n_{eM} \geq n^* \geq n_m$ for the risk-neutral case. However, Figures 2 shows that for the risk-averse case the relationship $n_{eM} = n_m > n^*$ is still possible for the relatively smaller values of R .

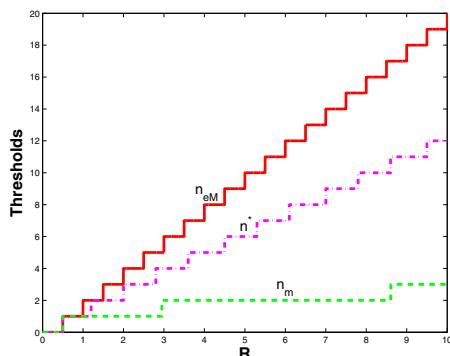


Fig. 1. Comparisons of thresholds with different partial information when $m = 0$

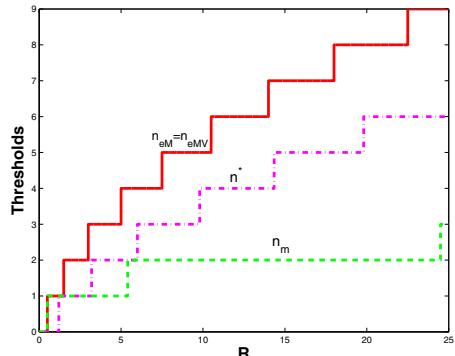


Fig. 2. Comparisons of thresholds with different partial information when $m = 1$

Acknowledgements. The authors would like to thank the support from the National Natural Science Foundation of China (No. 71101124), the Humanity and Social Science Foundation of Ministry of Education of China (No. 12YJC790101), the Natural Science Foundation of Hebei Province (No. G2012203068), the Scientific Research Project of Hebei Education Department (No. Z2011147, 2011120), and the Scientific and Technological Research Program of Qinhuangdao (No. 201101A550).

References

- [1] Guo, P., Zipkin, P.: The effect of the availability of waiting-time information on a balking queue. *European Journal of Operational Research* 198, 199–209 (2009)
- [2] Armony, M., Shimkin, N., Whitt, W.: The impact of delay announcements in many-server queues with abandonment. *Operations Research* 57, 66–81 (2009)
- [3] Sun, W., Guo, P., Tian, N.: Equilibrium threshold strategies in observable queueing systems with setup/closedown times. *Central European Journal of Operations Research* 18, 241–268 (2010)
- [4] Guo, P., Sun, W., Wang, Y.: Equilibrium and optimal strategies to join a queue with partial information on service times. *European Journal of Operational Research* 214, 284–297 (2011)
- [5] Kapur, J.N., Kesavan, H.K.: Entropy optimization principles with applications. Academic Press, Boston (1992)
- [6] Hassin, R., Haviv, M.: Equilibrium Behavior in Queueing Systems: To Queue or not to Queue. Kluwer Academic Publishers, Boston (2003)
- [7] Naor, P.: The regulation of queue size by levying tolls. *Econometrica* 37, 15–24 (1969)

Fault Diagnosis Based on Low-Frequency Acceleration Signals towards Large-Scale Recreational Facilities

Gen Li, Yu-tian Zhu^{*}, and Shuan-zhong Wang

College of Mechanical Engineering, Tongji University, 201804, Shanghai, P.R. China
yutianzhu@mail.tongji.edu.cn, reagan01@163.com

Abstract. Acceleration can be used as the parameter of fault diagnosis of large-scale recreational facilities. For it can not only reflect on the overall running condition of recreational facilities in the macroscopic view, but also analyze the impact in the process of running in the microscopic view. Firstly, evolutional wavelet based threshold denoising algorithms was used to deal with the acceleration signal and six characteristic values of the signal were extracted. Then the BP neural network was used for fault diagnosis. Finally the results were compared with FFT analysis results. The simulation was made by Matlab. The conclusion is that the results of fault diagnosis are reliable.

Keywords: fault diagnosis, neural network, acceleration signal, threshold denoising, wavelet analysis.

1 Introduction

The increasing accidents of large-scale recreational facilities occurred at home and abroad have drawn the attention of the security status of them. Hence, China has promulgated a series of laws, rules and regulations, technical specifications and technical standards on the safety of recreational facilities [1-3]. “Outline of the special equipment safety development strategies” which is implemented in January 2010, has made much clearer technical requirements for safety testing and monitoring. However, the existing test system at home and abroad is mainly qualified for static indicators of test such as the soundness of safety devices, structural parts for cracks, mechanical and electrical products, and it focuses on the special indicators on subsystems or components transmission mechanism or chain of a roller coaster. The effective means of real-time monitoring the state of the overall movement of recreational facilities is lacked.

The requirements and limitation about acceleration are introduced in "GB8408-2008 Amusement Device Safety Code". Acceleration signal can not only reflect on the overall running condition of recreational facilities in the macroscopic view, but also analyze the impact in the process of running in the microscopic view. Acceleration signal is often accompanied by noise and interference and need to denoise. Meanwhile, in order to carry out the fault identification and diagnosis effectively, the signal should be processed and extracted the required characteristic values. Therefore, according to the characteristics of the acceleration signal of large-scale recreational facilities,

^{*} Corresponding author.

wavelet based threshold denoising algorithms is used to deal with the signal, and six characteristic values of the signal are extracted. Then BP neural network is used for fault diagnosis and Matlab is used for simulation.

2 Characteristics of Low-Frequency Acceleration Signal

Low-frequency acceleration signal is a deterministic signal and can reflect the running track and status of the recreational facilities. The vibration frequency of the signal is very low, typically below 1Hz. The signal is very sensitive to changes. It should be taken seriously even if there is a small deviation between the real signal and the standard signal. What's more, there is evident difference between the low-frequency signal and white noise signal. So it is easier to filter out white noise signal.

3 Selecting the Characteristic Values

The principle to select the characteristic values is to reflect on the change of wave form. The six kinds of fundamental changes of wave form are shown in Fig. 1 (a) - (e).

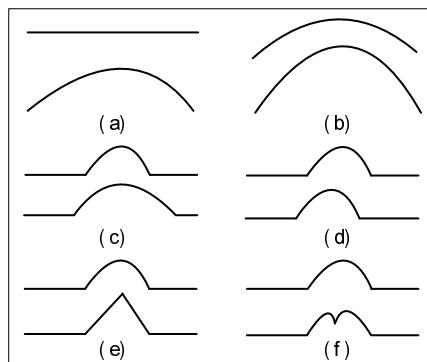


Fig. 1. Six kinds of fundamental changes of wave form

- (a) The change between emerging and vanishing; (b) The change of the amplitude;
- (c) The change of the length in the time domain; (d) The wave delays or advances in the time domain;
- (e) The wave form change to be sharp or gentle; (f) The peaks of increasing or reducing.

The peak indicators and the valid values can distinguish the change (a); The peak indicators, the valid values and the center of gravity of spectrum can distinguish change (b); The valid values and the magnitude margin can distinguish the change (c); The valid values and the center of gravity of spectrum can distinguish the change (d); The valid values and the kurtosis values can distinguish the change (e); The peak indicators and the pulse indicators can distinguish the change (f). To sum up, the peak indicators, the valid values, the magnitude margin, the kurtosis values, the pulse indicators and the center of gravity of spectrum are the selected characteristic values.

The fundamental changes of wave form can be reflected on by analyzing the six characteristic values.

4 Signal Processing

Firstly, a standard low-frequency acceleration signal was constructed. The signal was consisted of any elementary function by any addition and subtraction. Polynomial fitting method was used. We define the values of the coordinates of 100 points. After times of verification, we found the signal was the best when the order was 24. The standard signal curve is shown in Fig. 2 (a).

Then, a fault low-frequency acceleration signal was constructed. For the low-frequency acceleration signal is a deterministic signal, the signal can be segmented to process, which can improve the recognition accuracy of the characteristic values. The standard acceleration signal was divided into four sections. The first section was done nothing to change. And the second, the third and the fourth section was changed. These changes can reflect the six kinds of fundamental changes introduced in chapter 4. The fault signal curve is shown in Fig. 2 (b).

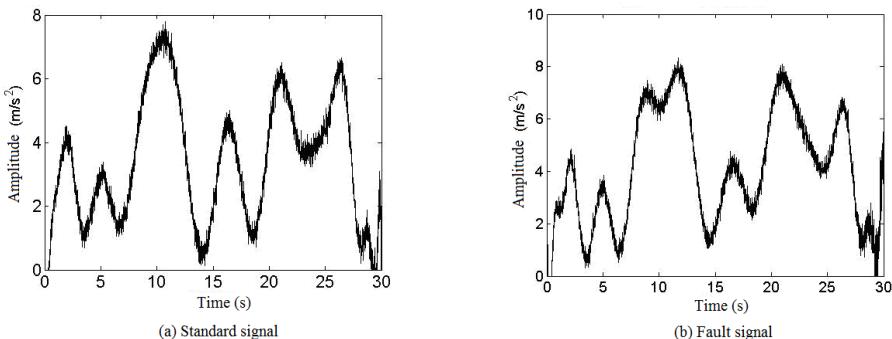


Fig. 2. Standard signal and fault signal

Low-frequency acceleration signal was denoised by using evolutional wavelet based threshold denoising algorithms. The principle is to filter out frequency band of noise or extract the needed frequency band of signal by the wavelet transform or wavelet packet transform.

Define $d_{j,k}$ as wavelet transform coefficient, $d'_{j,k}$ as estimated wavelet coefficient, j as wavelet decomposed scale, $k \in Z$, t as threshold, $t = \sigma\sqrt{2\log(N_j)}$, and $t > 0$.

The expression of the function of evolutional wavelet based threshold denoising algorithms is:

$$d'_{j,k} = \begin{cases} d_{j,k}(u^{2n} - p^{2n})/u^{2n}, & |d_{j,k}| \geq t \\ 0, & |d_{j,k}| < t \end{cases} \quad (1)$$

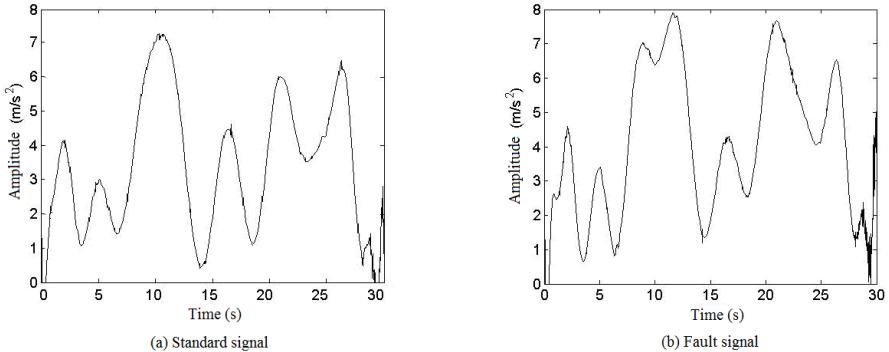


Fig. 3. Signals denoised by evolutional wavelet based threshold denoising algorithms

Where $u=(\exp(d_{i,k})+\exp(-d_{i,k}))/2$, $p=(\exp(t)+\exp(-t))/2$. u is arc hyperbolic function, and n is regulatory factor [4-6]. The results are shown in Fig. 2.

5 Fault Diagnosis

The fault diagnosis is completed by the BP neural network [7-9]. Firstly, the neural network should be trained. We divided the standard signal into four sections and extract the six characteristic values - the peak indicators, the valid values, the magnitude margin, the kurtosis values, the pulse indicators and the center of gravity of spectrum, of each section. The characteristic values were used as the input of the neural network, and then distribute the output of the neural network to each section. Train the neural network by setting the maximum number of training to 3000, the precision to 10^{-10} . Matlab 101The Matlab code of the training is as follows:

```

P=[firstm;P=[firstm;secondm;thirdm;forthm;]';
T=[1 0 0 0 0;1 0 0 0 0;1 0 0 0 0;0 1 0 0 0;0 1 0 0 0;
1 0 0 0 0;0 0 1 0 0;0 0 1 0 0;0 0 1 0 0;0 0 0 0 1 0;0 0 0 0 1
0;0 0 0 1 0];
net=newff([0 1;0 1;0 1;0 1;0 1;0
1],[13,5],{'tansig','purelin'},'trainlm');
net.trainParam.epochs=3000;
net.trainParam.goal=1e-10;
LP.lr=0.1;
net=train(net,P,T);

```

firstm; Where firstm, secondm, thirdm, forthm are arrays and consisted of the characteristic of the four sections. P is the input, and T is the output. net is defined as the neural network which the training is completed.

The error curve of the neural network is shown in Fig.4. The fluctuation of the error curve is low, which means the training is well and the accuracy of the neural network is high.

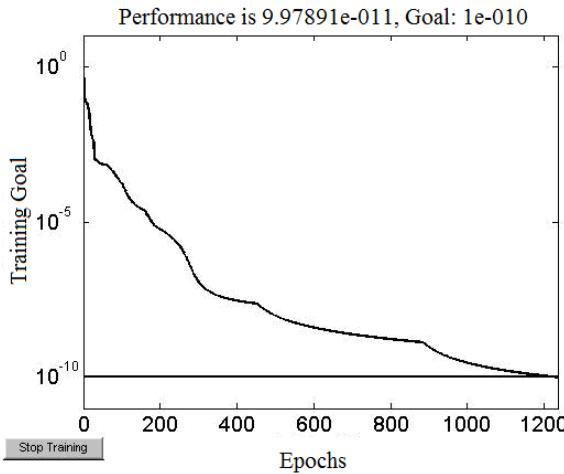


Fig. 4. The error curve of the neural network

The neural network which has completed the training can be used to diagnose the fault. We also divided the fault signal into 4 sections and extract the characteristic values of each section. The characteristic values were used as the input and the output was the result of the fault diagnosis.

To check on the result of the fault diagnosis, the FFT analysis of the standard signal and fault signal were made [10]. As is shown in Fig. 5, it is convenient to examine the differences. By contrast we can see, the result of the fault diagnosis by neural network is reliable.

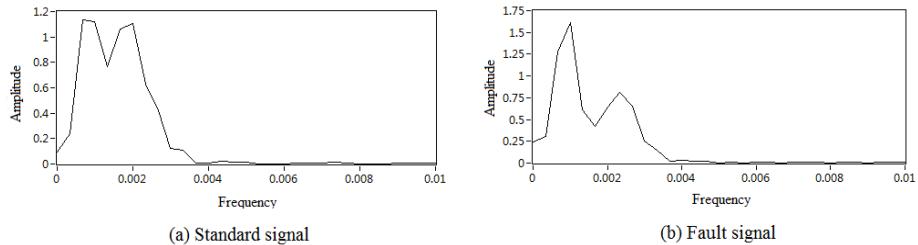


Fig. 5. Low-frequency acceleration signal frequency spectrum

6 Conclusion

Low-frequency acceleration signal is a deterministic signal. It can reflect on the overall running condition of recreational facilities in the macroscopic view, but also analyze the impact in the process of running in the microscopic view. It is an important parameter for the fault diagnosis of recreational facilities. In this paper, evolutional wavelet based threshold denoising algorithms was used to deal with the acceleration signal and six characteristic values of the signal are extracted.

Then the BP neural network was used for fault diagnosis. Finally the results were compared with FFT analysis results. The conclusion is that the results of fault diagnosis are reliable. And this method can be applied to the fault diagnosis system or monitoring and warning-forecasting system of the large-scale recreational facilities.

References

1. Liu, A., Yin, X., Liu, A., Yang, M., Chen, J., Zhou, S.: Establishment and Analysis of Safety Regulatory System for China's Recreation Facilities. *China Safety Science Journal* 19(5), 195–197 (2009)
2. Standards of People's Republic of China: Amusement Device Safety Code. GB, 8408–2008 (2008)
3. Li, G., Zhang, G., Ling, X., Huang, K., Qiang, T., Li, X.: Remote Security Monitoring and Warning - Forecasting System for Large - Scale Amusement Ride Based on Multi – Agent. *Computer Measurement & Control* 18(4), 824–826 (2010)
4. Jiang, K., Tang, X., Zhu, D.: Automobile Wheel Speed Signal Processing Based on Wavelet Algorithm of Improved Threshold. *Chinese Journal of Scientific Instrument* 31(4), 736–740 (2010)
5. Wang, S., Zhu, Y.: A Comparative Simulation Study of Evolutional Wavelet Based Denoising Algorithms. *Noise and Vibration Control* 32(1), 128–132 (2012)
6. Yuan, J., Zhang, D., Li, J.: Wavelet Denoising Algorithm Based on Improved Thresholding Method. *Journal of North China Electric Power University* 37(5), 92–97 (2010)
7. Bari, A., Bhasin, K., Karnawat, D.N.: Introduction to Neural Network and Improved Algorithm to Avoid Local Minima and Faster Convergence. In: Das, V.V. (ed.) CIIT 2011. CCIS, vol. 250, pp. 396–400. Springer, Heidelberg (2011)
8. Chen, T., Han, D., Au, F.T.K., Than, L.G.: Acceleration of Levenberg-Marquardt Training of Neural Networks with Variable Decay Rate. *IEEE Trans. on Neural Network* 3, 1873–1878 (2003)
9. Zhang, Z., Zhang, J., Ye, L., Zheng, Y.: Research on the Application of Neural Network in Diaphragm Icing Sensor Fault Diagnosis. In: Yu, W., He, H., Zhang, N. (eds.) ISNN 2009, Part I. LNCS, vol. 5551, pp. 589–600. Springer, Heidelberg (2009)
10. Zhao, C., He, M.: Short Time Fourier Analysis Based on the Special Frequency Component. *Automation of Electric Power Systems* 28(14), 41–44 (2004)

Research on Embedded-Based Wireless Authentication System

Wei Pan¹, Xiang Yang^{2,3,4}, Yuanyi Zhang⁵, and Zhe Guo¹

¹ School of Mechanical and Control Engineering, University of GuiLin Science and Technology, Guilin 514000, China

² School of Information Engineering, Wuhan University of Technology, Wuhan, 430070, China

³ School of Information Science & Engineering, Guilin University of Technology, Guilin, 541004, China

⁴ BoWen School of Managemet, Guilin University of Technology, Guilin, 541006, China

⁵ School of Architecture, Fuzhou University, Fuzhou, 350108, China

243546041@qq.com, yangx@glite.edu.cn, zyye@fzu.edu.cn,
guozhelong@sina.com

Abstract. Fingerprint identification is a high technology that extensively used in recent years. It proposed a new authentication system design based on fingerprint identification and 3G technology which uses S3C2440 as its microcontroller and MBF200 as its fingerprint sensor. With this system, certificates can be identified accurately and some cornhill robbery can be avoided. Also the authentication process is rapid and convenient There is no need to worry about the false testimony or certificates be stolen or lost. As a result, it helps to reduce some crimes and contributes a lot to establishing a harmonious society. Also a testing to the system was done do and the experiment results show its security and stability. This system is more advanced than traditional ones by using the latest technic and possesses good application value of popularization.

Keywords: Fingerprint identification, 3G technology, S3C2440, MBF200.

1 Introduction

Archaeologists have confirmed that fingerprint as a authentication tool has been used both in ancient Syria and China in roughly 6000 to 7000 B.C. Scientists find the reasons why fingerprint is always be used is that different people have their different fingerprint even twins [1].And the fingerprints will not changed. In other words, it is permanent and forever existed until people's death. As a result, it could be a excellent means of authentication. In recent years, with the development of the computer and Optical scanning technology, biological fingerprint identification technology is becoming mature gradually. From now on, biological identification technic is the most convenient and securest technology [2].

As the technology developed, some social problems appearing too. Electronic information crimes are ramification from modern high-technology development. In

some countries especially in developed ones, there were a large number of lawbreakers stealing a the identity information of others for the characteristics of the “phishing” and other new types of crimes [3]. The development of authentication system is in urgent demand especially in our modern harmonious society.

2 Design Program

At present, there are various biological identification technologies at home or abroad including facial recognition, signature recognition, iris recognition, retina recognition, fingerprint identification technology and voice recognition technology and so on. Several commonly used biometric authentication based on technical characteristics point of view, the technical solution based on fingerprint identification is far superior to the other five. Moreover, with the third generation mobile communication network (3G) services worldwide, the increasing popularity of mobile Internet bandwidth brought about by the increase in technology-driven force has greatly promoted the development of wireless mobile services. Therefore, this article will use 3G technology, fingerprint identification technology and embedded technology, proposing a new fingerprint-identification-based wireless authentication system. It provides an ideal solution for the problem of current social authentication and information security technology.

3 Design of Hardware Platform

This design of Embedded- Technology-Based Wireless Authentication System is mainly consisted of core control module, power module, memory module, fingerprint collection module, communication module and interaction and operation module. The structural framework of system is shown in Fig.1.

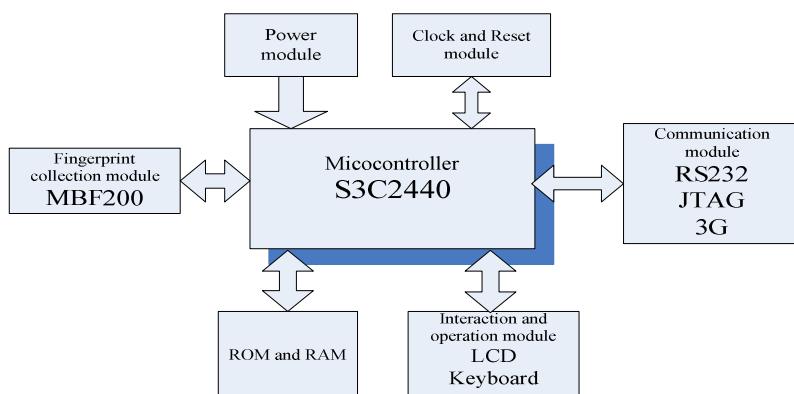


Fig. 1. The structural framework of system

The main controller of the core control module is S3C2440, which is 16/32-bit RISC microprocessor produced by the Samsung company. It was developed using the ARM920T core. It is mainly responsible for overall system control, and coordination of work between each module. All of integrated on-chip functions[4] can help us to reduce the system design cost and convenient for us to develop a product with low cost.

Power management module provided regulated power supply for the system's normal work. Fingerprint acquisition module used fingerprint sensor as the core, it can acquire the fingerprint image in the real-time under the control of the main controller, and transmit image data to the system processing. ROM is used to store embedded operating systems, applications, fingerprint recognition algorithm and the fingerprint characteristic data. RAM is used to extend the temporary data space implementation which needed by the program, the storage space mainly depends on the needs of fingerprint recognition algorithm. Communication module is responsible for the data exchange between systems and peripherals, and the communication functions between the server-side of Internet and the system. Interaction module included keyboard and LCD. The keyboard is responsible for receiving user's input information, and notification system to make the appropriate treatment. LCD used to displaying the user interface, and provide system message and the results to facilitate user actions and use.

During the whole process of system design, fingerprint acquisition module and the 3G communication module is the most crucial factors.

3.1 Fingerprint Acquisition Module

Fingerprint identification technology involves fingerprint image acquisition, fingerprint image processing, features extraction, data saving, features matching, and so on. So the performance of fingerprint sensor will directly affect the whole system. There are many fingerprint sensors in the market at present, such as MRB2000, FPS110, FPS200, TMS320VC559, and so on. Most of them have bad image quality, lower antistatic property, and poor environmental adaptability for different kinds of fingers. Therefore, we used a capacitive fingerprint sensor MBF200 which contains some characteristics such as the high performance, low power dissipation, low cost

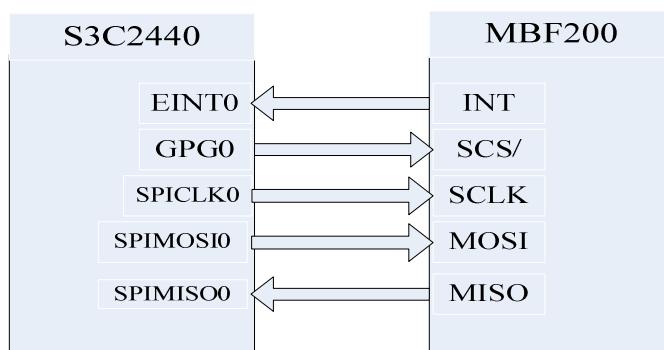


Fig. 2. The schematic diagram of S3C2440 and MBF200 interface

and so on. It has three interface modes: microprocessor interface modems interface mode and USB interface mode [5]. In this design, we use the SPI interface mode. The interface schematic diagram is simple as shown in Fig. 2.

3.2 3G Communication Module

3G communication interface also is an important part of this system. It realized the connection between the system and the server which was on the Internet (or LAN), so that to form a large common authentication system with the servers on the network.

The system used Huawei's latest EM770W UMTS M2M wireless module. The module was built-in TCP/IP protocol stack. It has many functions such as voice, text messages, data, phone book and supplementary services, FOTA and so on. The band it supported including UMTS 2100/1900/900/850 and GPRS/GSM1900/1800/900/850. EM770W has abundant interface resource, it provided a common Mini PCI Express interface. The interface signals were: power, PCM interface, a 2-way UART interface, a way of input and output audio signals, a way of USIM card and a road speed signal USB2.0 interface. At the same time, EM770W supported the standard AT command set and extended AT command set of Huawei, and it also supported the special USIM card business, and met ROHS environmental certification requirements [6].

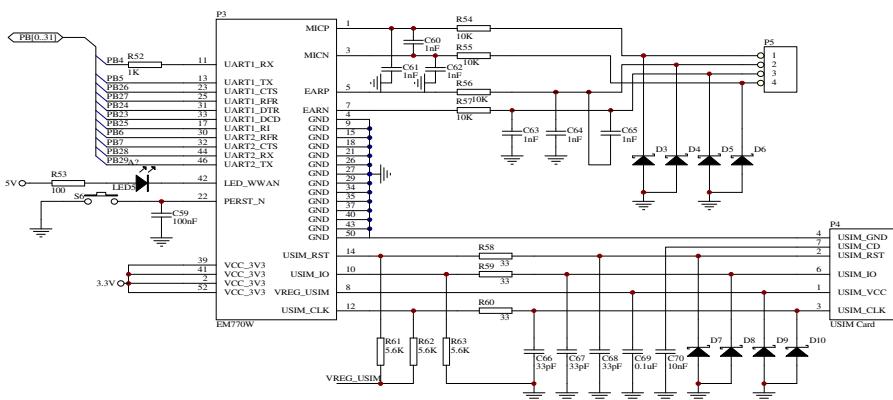


Fig. 3. The schematic diagram of 3G communication interface

The schematic diagram of 3G communication interface was shown in Fig.3. EM770W UMTS M2M was composed of UART interfaces, LED status indicator and RESET Interface, Audio Interface, USIM card interface and so on. Which, UART1 with flow control features to support all serial mode, and to support DATA service that allows users to dial the UART1 out from the PPP, and the data services operations. UART2 supports 5-wire serial mode, and sending and receiving DIAG command. The RX pin of UART1 connected with the PB4 pin of S3C2440 through a 1K resistor, in order to avoid the level caused by different I/O port damage. EM770W provides a LED status indication interface “LED_WWAN”, this pin can be

configured as a programmable current source to drive an external indicator. Of course, the R53 of Fig.4 can be replaced by a variable resistor to adjust the LED's brightness. “PERST_N” was the reset pin of EM770W, it can achieve EM770W hard reset by directly pin down this pin for 100ms. Audio interface is the audio input and output interface of EM770W, it was mainly used to record and enter some authentication tone. As the audio input and output communication is fully differential signal, and the RF performance was very well, so it needn't to extend the other external audio amplifier. The rate of USIM card interface is generally 3.25MHz, so the USIM Card should be as close as possible to near EM770W, to avoid causing serious distortion to affect communication signals. In addition, added a 0.1uF capacitor to pin USIM_VCC, and added three 33pF capacitors to the pins USIM_CLK, USIM_IO and USIM_RESET with the ground, can filter out the interference of antenna signal.

4 Design of Software System

Linux2.6 is embedded to the software design and modular programming is used to guarantee its good portability and good expansibility. The bottom is drive layer. It is composed of various device drivers, and responsible for providing service interface to the upper layer. The service interface has a uniform data format, and the data format will not be changed no matter whether hardware equipments change or not.

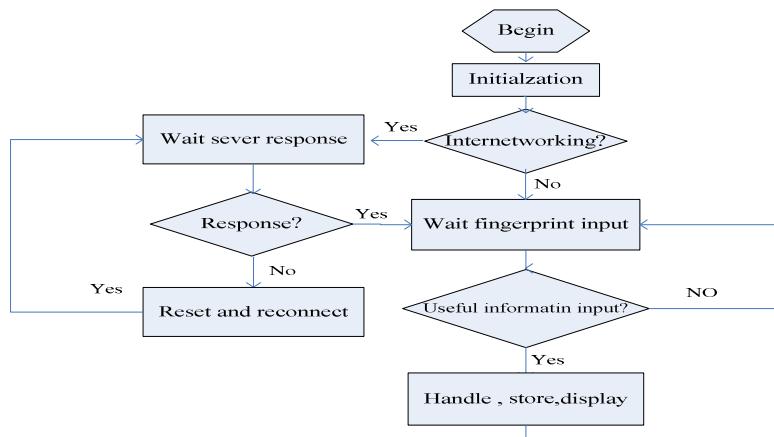


Fig. 4. Main software structure of the system

As the fingerprint sensor MBP200 communicated with the main controller S3C2440 through Serial Peripheral Interface (SPI). So the system will loading SPI drivers in the beginning. First, it defined name and hardware resource of MBF200 through structure body “struct_device”:

```
struct spi_device{
    const char *name;
    u32 id;
    struct device dev;
    u32 num_resources;
    struct resource *resource;
};
```

Then, spi_device tell system kernel to load the device, through enroll structure body “struct device dev”. After the FPC1011C driver has loaded successfully, it will produce two folders: devices and drivers, in each directory below the bus. The main software structure of the system is shown as in Fig. 4.

5 System Testing

For testing the recognition rate of system, we acquired the fingerprint information through MBF200, then set and stored the start address and length of fingerprint data under the CCS interface, which as the fingerprint template. Secondly input the basic information of users, including name, sex, nationality, ID number, date of birth, home address, photo and so on, testing in the stand-alone mode. Then, obtained the fingerprint once again through MBF200, and compared with the fingerprint template which stored on the database in 3G online mode. When the matching result is right, the system will display the identity information of the user automatically, otherwise, it will display the warning information to suggest that the user is not existed. Fig. 5 shows the authentication interface of this system.

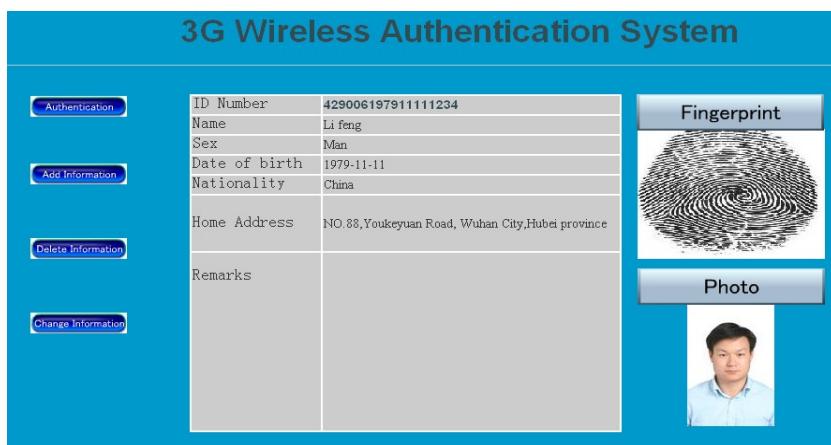


Fig. 5. The authentication interface of system

After doing many tests to different people. The system works stably and accurately. It has strong adaptability whether to dry or wet fingers. During the test, FAR was 0.007, and the FRR (False Rejection Rate) was 0.0002, and the total time of each match are within 2.0s. The experiment results show that this system has the superiority that traditional authentication system were not. The system provides an ideal solution for authentication and information security problems, which will contributes to establishing a harmonious society.

References

1. Research of fingerprint Authentication system, <http://bbs.81tech.com/forum>
2. Zhao, J., Li, H.: Desin and implementation of fingerprint recognition system. Academic Research of Shijiazhuang Lionful Vocational College (2) (2007)
3. Yang, C.: The international development and its implication of identity crime legislation. Political Science and Law (8), 35–39 (2008)
4. S3c2440a 32-Bit Cmos Microcontroller User's Manual Revision 1.Pdf
5. Wang, Q., Zhu, Z.-J.: Fingerprint Acquisition System Based on TMS320VC5402. Communications Technology (7) (2010)
6. Huawei_Em770w_Datasheet_V3_0.pdf
7. Chen, S., Lu, J.: Design of storing compartment system based on fingerprint identification. Modern Manufacturing Engineering (5), 118–120 (2008)
8. World Leader in Fingerprint Authentication Technology,
<http://www.veridicom.com/>
9. Yang, X., Wang, H., Pan, W., Zhang, Y.: An Embedded- Technology-Based Web Server Design of Digital Home. Advanced Materials Research (219-220), 820–823 (2011)

Data Reliability Analysis for Flooding Transmission of Wireless Sensor Networks

Cao Yuan-Yuan and Feng Hai-Lin

College of Science, Xidian University, Shannxi, 710071, China

Abstract. This paper proposes a reliability strategy to solve the problems of energy greatly damaged and low data reliability while using traditional flooding routing in wireless sensor networks. The strategy combines directional flooding data transmission and network coding. By limiting transmission direction and selecting part of nodes in the directional transmission range for encoding, the strategy can improve data reliability and reduces energy consumption. Theoretical analysis and simulation results show the proposed strategy is effective.

Keywords: Wireless Sensor Networks, Network Coding, Reliability, Encoding Node.

1 Introduction

Flooding leads to node energy badly damaged and the network lifetime has been affected. In order to overcome the defects of the traditional flooding routing, restricted flooding routing strategies have been proposed. This paper propose a reliability strategy for flooding data transmission based on network coding, aiming at reduce energy consumption by limiting transmission range and in the data transmission, selecting some encoding nodes to improve data reliability [6,7,8], the strategy makes full use of network coding to increase the transmission reliability while reducing the complexity of network caused by network coding.

2 Network Model

Assume N sensor nodes are deployed in a two dimensional area with a density ρ , nodes and wireless links may be random failure, e and ε are failure rates respectively. This model use random linear network coding scheme [9], which does not need to determine the coefficients of coding, very suitable for network node or link time-varying WSNs. We give the following definitions:

Definition 1[10]: Data transmission reliability R based on network coding is the probability that sink node receives all the packets successfully after a transmission.

Definition 2: Encoding nodes: In the coding scheme, a node v is called a encoding node if it encodes one or more input packets and forward along the channel.

3 Data Transmission Based on Network Coding

3.1 Data Transmission Scheme

This paper limited transmission range when the node sends a packet. As show in fig.1, source node S sends packets to destination D , if S only transmit packets to nodes A, B, C , which are on the direction of D , the packets need less hops and save the node energy.

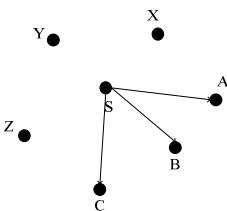


Fig. 1. Directional sends

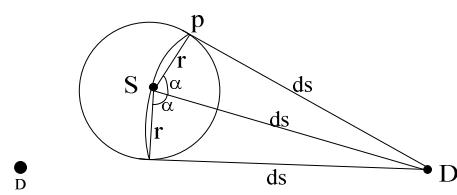


Fig. 2. Forwarding angle

As show in fig. 2, d_i is the distance between neighbor node i and D . The radius of circle D is d_s , interseting with circle S at the point p . At ΔSPD , according to

the cosine theorem, $\cos \alpha = \frac{d_s^2 + r^2 - d_s^2}{2rd_s}$, simplified and calculated to

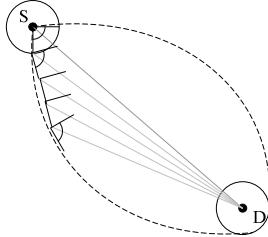
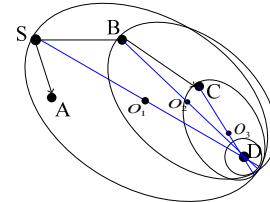
$\alpha = \arccos \frac{r}{2d_s}$. The original flooding transmission is limited and transmitted within

the range of line $|SD|$ with the offset angle α . Without loss of generality, the

transmission range of neighbor node i is $\alpha = \arccos \frac{r}{2d_i}$.

For each node, whether it could be an encoding node depends on the number of its neighbors. If we select the nodes within scope α on the border for data transmission, as shown in fig. 3, the forwarding area looks like a ellipse. Therefore, this paper specifies ellipse as the forwarding area, adjust the number of neighbors through dynamic control of the data forwarding area. The elliptic long axial is the distance between the source node and destination node, while the short axial length depends on actual needs, which can dynamically change according to the node density ρ . The greater the density, the more the number of neighbors of a node.

As shown in Fig.4, deploy nodes with a density ρ , source node S sends packets to destination node D , circle D is inscribing with ellipse O_1 , elliptic long axial is α , nodes within the ellipse O_1 to be able to receive packets sent to S . When a packet is sent to the circle D , it can choose next hop node to terminate packet transmission.

**Fig. 3.****Fig. 4.** Flood transmission based on ellipse

The area of the ellipse is $S_{ellipse} = \pi ab$. We must ensure that network coding in the new mode of transmission smooth, that is to ensure the number of neighbor nodes within a certain range. Elliptic area $S_{ellipse}$ and density ρ into inverse proportion,

$$\rho \propto \frac{1}{S_{ellipse}}, \text{ then } \rho \propto 1/b. \text{ A node in the ellipse has the coordinates of } (x_i, y_i), \text{ if } \frac{x_i^2}{a^2} + \frac{y_i^2}{b^2} \leq 1, \text{ node can receive packets.}$$

3.2 Encoding Node Selection

Paper [11] proposes that under the known of the local information of network topology and the same loss rate e of channel, the node i whose degree is x_i can receive the packets from neighbors is proportional to x_i^α . It means that when the x_i is higher, the more packets received from the neighbors.

The benefit of network coding to improve reliability is to add more redundancy packets in the transmission. Therefore, in this case, the encoding operation can be used. Since the neighbors also receive packets from their neighbors, we determine the probability as follows.

Definition 3 (Probability p_i): The probability of node i in the network to become encoding node is p_i . Where x_j is the degree of node i .

$$p_i = 1 - \frac{x_i}{\sum_j x_j} \quad (1)$$

The first special case is the isolate node in the network; we define these nodes as general nodes. The probability of these nodes to become encoding nodes is zero.

The second case is the node i follows the formulate, Where k is the degree of i .

$$x_i = \sum_{i=1}^k x_i \quad (2)$$

As shown in Fig. 6, node 5,6 follow equation (2).According to the definition 3, the probability that these nodes become encoding nodes is zero, so we call the nodes in Fig.6 belongs to the set of isolate nodes. In this paper, we denote the probability of any node in these set is zero.

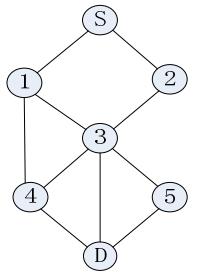


Fig. 5. Node topology

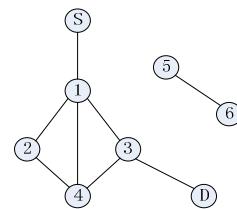


Fig. 6.

We define P_i as follows

$$P_i = \begin{cases} 1, & \text{when } i \text{ is source node} \\ 0, & \text{when } i \text{ is sink node, isolate nodes} \\ p_i, & \text{otherwise} \end{cases}$$

$N(u)$ is the set of neighbor nodes of u ,because of the limits of the transmission range, the number of neighbor nodes of u reduced as $N(u)^*$, $N(u)^* = \{v_u^1, v_u^2, \dots, v_u^{|N(u)|}\}$,where $N(u)^* \subseteq N(u)$. Assume we choose k encoding node $V_c = \{v_c^1, v_c^2, \dots, v_c^k\}$, source node $S \notin V_c$.

To ensure the superiority of network coding, we set a timer on each node,the size is T .The nodes receive the packets within T ,and encode or forward the packets when the timer equal to zero. After sending the packets, the timer is reset to T . Data transmission strategy as follows:

- Step1: selected source node S and destination node D ;
- Step2: through the angle $\alpha = \arccos \frac{r}{2d_u}$ and the limited of ellipse, to determine the neighbor set $N(u)^*$ of u ;

- Step3: compute the P_i of this node, and determine the encoding node set V_c ;
- Step4: if $u \in V_c \cup \{S\}$, encoding the received packets of node u and forward encoding packets to its neighbors $N(u)^*$, node u encode the redundancy packets according network coding strategy; if $u \notin V_c \cup \{S\}$, only forward the packets to its neighbors $|N(u)^*|$;
- Step5: traverse $N(u)^* = \{v_u^1, v_u^2, \dots, v_u^{|N(u)^*|}\}$, if $D \in N(u)^*$, successfully received packets, stop forwarding; otherwise go to step6;
- Step6: take a node in the set of $|N(u)^*|$ as the new forwarder u , go to step4.

4 Reliability Analysis of Data Transmission

Because of the limited transmission range, the number of neighbor nodes of u reduced to $|N(u)^*|$, $|N(u)^*| = |N(u)| \times \frac{2\alpha}{2\pi} = |N(u)| \times \frac{\alpha}{\pi}$. There are $|N(u)^*|$ neighbor nodes send $|N(u)^*|$ packets. Suppose in addition to the source node, selected m encoding nodes. According to their distance from the source node label ascending. Source node to the first encoding node requires n_0 hop, the first encoding to the second encoding node requires n_1 hop, ..., the $m-1$ encoding node to the m encoding node requires n_{m-1} hop, the m to destination node requires n_m hop.

Suppose the source node in the transmission range has $|N(S)^*|$ neighbors, need to send M packets, encode the M packets and forward $M'(M' > M)$ encoding packets to its neighbors. In order to ensure the destination node can successfully decoded, destination node and all encoding nodes are received at least M linearly independent packets.

Consider link and node failure, link and node failures are independent, channel transfer rate is $q = 1 - e$, a node does not failure as the probability $\xi = 1 - \varepsilon$. There are $|N(S)^*|$ neighbor nodes send $|N(S)^*|$ packets. When the $|N(S)^*|$ at least one neighbor node receives the encoding packet is sent successfully.

The probability of one hop successful transmission is $(1 - (1 - q\xi)^{|N(S)^*|})$.

Therefore, The probability of n_i ($0 \leq i \leq m$) hops successful transmission is

$$q_i = (1 - (1 - q\xi)^{|N(S)^*|})^{n_i} \quad (3)$$

Suppose an average of n_i hops from the i encoding node to the $i+1$ encoding node. The probability of the $i+1$ encoding node at least received M linearly independent packets is :

$$Q_i = \sum_{k=M}^{M'} C_{M'}^k q_i^k (1 - q_i)^{N-k} \quad (4)$$

Therefore, the data transmission of reliability from source node to destination node is:

$$Q = \prod_{i=0}^m Q_i = \prod_{i=0}^m \left(\sum_{k=M}^{M'} C_M^k q_i^k (1-q_i)^{N-k} \right) \quad (5)$$

5 Analysis of Energy Consumption for Data Transmission

Suppose β is the energy that node transmits a packet required for, source node sends $|N(S)|M$ packets. After n_i ($0 \leq i \leq m$) hops successful transmission, the energy required for receiving packets: $E_{rev} = n_i |N(S)|MQ_i\beta$. The instability of wireless links, a certain amount of packet loss, after j -hop lost packets, the energy consumption is,

$$E_i = E_{rev} + E_{los} = n_i |N(S)|MQ_i\beta + \sum_{j=1}^{n_i} (|N(S)|M \times \beta \times j \times q_i^{j-1} (1-q_i)) \quad (6)$$

γ is the energy consumption of the node for encode or decode packets, for the M' encoded packet,

$$E'_i = E_i + \gamma = E_{rev} + \sum_{j=1}^{n_i} E_{los} + \gamma \quad (0 \leq i < m) \quad (7)$$

$$E'_m = E_m + 2\gamma = n_m |N(S)|M'Q_m\beta + \sum_{j=1}^{n_m} (|N(S)|M' \times \beta \times j \times q_m^{j-1} (1-q_m)) + 2\gamma \quad (8)$$

Total energy consumption is:

$$E = \sum_{i=0}^{m-1} E'_i + E'_m = \left(\sum_{i=0}^m n_i |N(S)|M'Q_i\beta \right) + \sum_{i=0}^m \sum_{j=1}^{n_i} (|N(S)|M'\beta j q_m^{j-1} (1-q_m)) + (m+2)\gamma \quad (9)$$

6 Simulation

In this paper, under Matlab7.0 environment. The experiment model is in the region of $100m \times 100m$ with 100 nodes deployed randomly. We simulate the case of $r = 20$.

The number of encoding nodes Fig. 6.1 and Fig 6.2 show the relationship between the number of encoding nodes and network topology density. When $2b \geq 20m$, encoding nodes decreases with the decrease of b . When the network

topology as the node density increases and becomes complex, the encoding nodes also reduces.

Reliability $2b = 50m$, source node encoded 3 packets into 5 encoding packets. As show in fig6.3, network coding transmission reliability is superior to the traditional one. The gap is obvious after the transfer rate more than 0.55. Fig6.4 shows that after the transfer rate more than 0.55, the gap of the reliability is not obvious between traditional transmission and elliptic transmission.

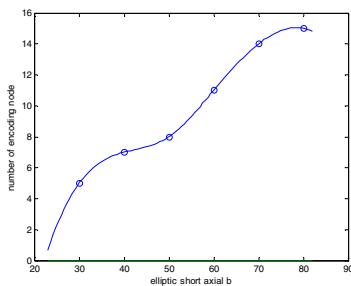


Fig. 6.1. Encoded node with the chang of b

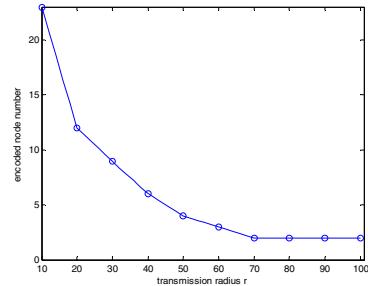


Fig. 6.2. Encoded node with the chang of r

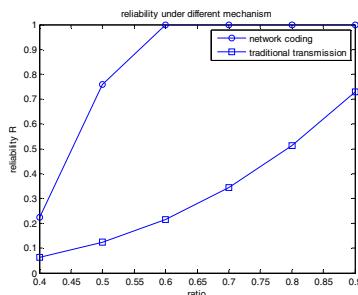


Fig. 6.3. Reliability under different mechanism

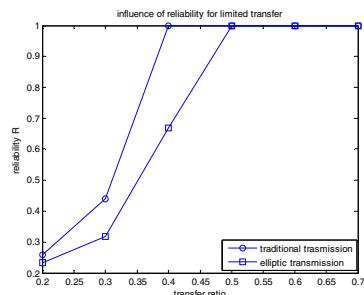


Fig. 6.4. Reliability for limited transfer

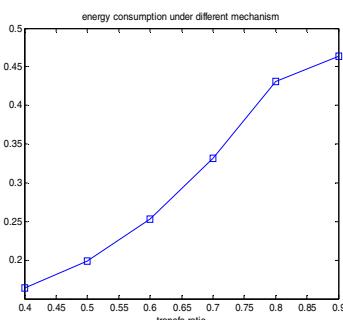


Fig. 6.5. Ratio of energy consumption

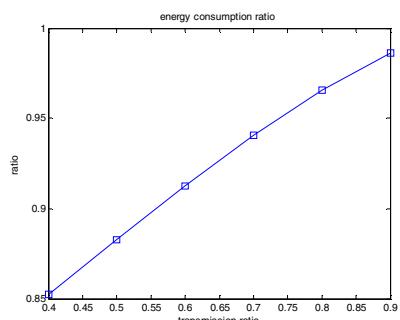


Fig. 6.6. Ratio of energy for ellipse and not ellipse

Energy consumption Fig6.5 and fig6.6 show that different transfer rate of the ratio of energy consumption. Fig6.5 shows that the ratio increased when the transfer rate is high, ensure the successful transmission at the same time, some energy will be spent. Fig6.6 is the same as fig6.5.

7 Conclusion

This paper proposes a reliability strategy for flooding data transmission based on network coding. In this strategy, we improve the reliability of data transmission, and reduce the number of encoding nodes, thereby reducing energy consumption of nodes. The transmission is under the improved flooding routing, we get the conclusion that R of using this coding strategy is higher than traditional transmission.

Acknowledgments. The author would like to thank the anonymous reviewers for their insightful comments and suggestions for improving the paper. This research was supported by the National Natural Science Foundation of China under Grant No.60874085 and No.60974082.

References

1. Rout, R.R., Ghosh, S.K., Chakrabarti, S.: A Network Coding based Probabilistic Routing Scheme for Wireless Sensor Network. In: 2010 Sixth International Conference on Wireless Communication and Sensor Networks (WCSN), pp. 15–19 (2010)
2. Tang, D.-P., Luo, J.-Y.: The Research on Routing Algorithms Based on Wireless Sensor Network. Microcomputer Information, 175–179 (2009)
3. Zhang, J., Lin, Y.-P., Li, C., et al.: An Angle-Area-Based Flood Routing Algorithm for Sensor Networks. Computer Engineering & Science, 55–77 (2005)
4. Ahlswede, R., Ning, C., Li, S., et al.: Network information flow. IEEE Transactions on Information Theory, 1204–1216 (2000)
5. Yang, Y.-X.: Network coding theory and techniques, pp. 1–4. National Defense Industry Press, Beijing (2009)
6. Lin, S., Costello, D.J., Miller, M.J.: Automatic-repeat-request error-control schemes. IEEE Communications Magazine, 5–17 (1984)
7. Ufuk, D., Ozlem, A.: Raptor versus Reed Solomon forward error correction codes. In: Proceedings of ISCN 2006: 7th International Symposium on Computer Networks, pp. 264–269 (2006)
8. Deb, B., Bhatnagar, S., Nath, B.: Reinform: Reliable Information Forwarding using Multiple Paths in Sensor Networks. In: Proceedings of 28th Annual IEEE Conference on Local Computer Networks-LCN, pp. 406–415 (2003)
9. Ho, T., Muriel, M., Shi, J., et al.: On Randomized Network Coding. In: Proceedings of 41st Annual Allerton Conference on Communication Control and Computing, pp. 1–10 (2003)
10. Hao, W., Chuang, L., Fengyuan, R., Yao, Y., Xiaomeng, H.: Retransmission or Redundancy: Transmission Reliability in Wireless Sensor Networks. In: IEEE International Conference on Mobile Adhc., pp. 222–228 (2008)
11. Yin, C.-Y., Wang, B.-H., Wang, W.-X., et al.: Efficient routing on scale-free networks based on local information. Physics Letters A, 220–224 (2009)

Effective Management and Utilization of Mobile Learning Based on Wireless Campus Network

Shujie Du, Zhuan Zhou, Lanfang Liu, Ling Jiao, and Bo Wu

Computing Centre, Ocean University of China 266100 Qingdao, Shandong, China
{qingdaobeach, zhou_zhuan}@gmail.com,
{11fang0710, jamison, zs_ouc}@ouc.edu.cn

Abstract. Combination of wireless campus network and portable computers makes personalities expand anywhere. This paper analyzes the new generation of the wireless mainstream application technology 802.11n, discusses the construction planning and design of the wireless campus network which supports mobile learning. Intelligent mobile terminals are adapted to independent or interactive e-learning on network. Here gives design architecture with concern to support various kinds of mobile learning, puts up with some improved opinions dealing with the key technology on management and security.

Keywords: wireless access, 802.11n, IPv6, mesh, QoS.

1 Introduction

The continuous development and widespread application of the mobile technology and digital technology prompt the development of wireless campus network construction and transformation of educational models. Relying on the mobile communication network, business such as the teaching, management and services becomes more convenient. The changes of teaching methods and high bandwidth business applications present more requests to wireless campus network construction, effective construction of wireless campus network to support mobile learning is imminent.

2 Planning for Wireless Campus Network Supporting Effective Mobile Learning

2.1 Wireless Campus Network Take the Mainstream Application Technology by 802.11n

With the release of the 802.11n technology standard, to use the representative of advanced access technology for wireless network- 802.11n protocol to construct a new generation of wireless campus network has become a trend. As a new generation

of WI-FI standard, the 802.11n protocol uses 2.4GHz band and 5GHz band, coverage improved by 20%-30% compared to the traditional wireless technology. It has good bearing capacity. 802.11n transmission bandwidth is 5 times of the traditional one. Theoretical rates up to a maximum of 500Mbps. the actual rate can be more than 100Mbps; 802.11n also provides concurrent dual-band operation, so as to provide more channel capacity for broadband multimedia applications, meet the needs for high bandwidth, high quality mobile learning service [1].

Table 1. Comparison of IEEE 802.11 wireless communication technologies

	802.11a	802.11b	802.11g	802.11n
Transmission Rate	54Mbps	11Mbps	54Mbps	+100Mbps
Band	5GHz	2.4GHz	2.4GHz	2.4~5GHz
Transmission Distance	70m	100m	100m	+100m
IEEE Authentication	1999	1999	2003	2007

In 802.11n, MIMO (Multiple Input Multiple Output) is the core technology, using a plurality of antennas to improve the quality of transmission: the receiving antenna also can receive the signal after reflection or scattering except for the maximum signal. It's suitable for the conditions several people use a wireless network, for example: in the classroom, wireless network user is using a separate wireless network node, so other users can simultaneously use the node. In the traditional wireless networks, if several users use a node simultaneously, the network performance is weakened. As a result of MIMO, a separate antenna can also be assigned to multiple users to use, and its speed had no noticeable influence.

In actual campus environment, there are a lot of 802.11b/g wireless network card users. In 802.11n mode, It is downwardly compatible with 802.11b/g or 802.11a, which will affect the overall performance of 802.11n equipment. Therefore in the 802.11n device through adjusting RF configuration, two radio frequency to provide user access. For example, a way of working in the 5 GHz band is configured as a 802.11n model, specifically for 802.11n card users to provide high speed access; the other way at 2.4G band is configured as mixed compatibility mode, to ensure that the traditional 802.11g customer's normal access, for different users respectively provide the best adaptive internet access.

2.2 Wireless Campus Network Mode

Autonomous Mode. Autonomous mode is also known as the Fat AP networking mode, AP devices work as separate individual, gathering the WLAN physical layer, user data encryption, user authentication, QoS, network management, roaming technology and other application layer functions in a AP, receiving and sending data through wireless mode with users.

Centralized Control Type AC+Fit AP Networking Mode. In this kind of wireless network framework, the access controller AC (wireless switch) is required for the unified configuration management. Fit AP itself does not have any software and

configuration, AC achieved radio frequency monitoring, data flow management, security certification, QoS, access control, load balance and the AP management control functions. AP only achieves 802.11's empty function, complete radio transceiver task. The essence of using the AC+Fit AP mode is to reduce the complexity of the AP, to transfer the network intelligence to upper control equipments, in order to solve the problems that happened in the Fat AP network structure mode.

All data must be forwarded to AC for processing when using AC centralized forwarding mode. It's easy for centralized planning and maintenance in this way. Also it's flexible for configuration. The disadvantage is that AP can become the bottleneck of wireless network easily.

Mesh Networking Mode. Wireless Mesh Networks (WMN), also be known as wireless mesh network or wireless multiple networks, which is mobile self organized network that can be called as a multi-hop combining the fixed and mobile node together using wireless network [2]. In wireless Mesh networks, any wireless device node can act the AP and router at the same time, each node in the network can send and receive signals, each node can communicate with one or more peer nodes directly.

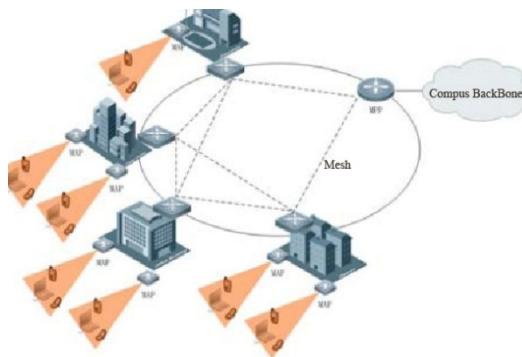


Fig. 1. Wireless mesh campus network

Wireless campus network has the following characteristics: The first one is the campus wireless network not only has huge geographical range, large number of users, but also the large amount of communication, as in mobile learning, students will use multimedia much more often. The second one is the high requirements for the network coverage, network must be able to realize seamless roaming among the indoor, outdoor, auditorium, dormitory, libraries, and other public places. The third one is the importance of load balance, because the students often have to concentrate, when many students in a place using the network at the same time there may occur the phenomenon of traffic congestion. To solve these problems, the traditional method of installing AP in the interior in high density, while AP quantity is little in outdoor installation. But due to the campus network user demand is changing greatly, add new AP or adjust AP deployed position is often happen [3,4], which increase a lot of cost. In wireless Mesh networks, AP increase or adjustment become easier, configuration

and installation are more flexible, and cost are lower, especially for the area that the users often need change access points, the wireless Mesh technology multichip structure and flexible configuration will be very beneficial to network topology structure adjustment and upgrade. To use Mesh network, deploy Mesh nodes correctly, better finish the open area of wide coverage, and finish the open area's wide coverage. not only be easy to realize network structure upgrading and adjustment, but also allow the network users share the bandwidth, eliminating the bottleneck in current single hop network, realizing the network load dynamic balance, in the wireless management capabilities, it has the advantages of centralized control, in safety management aspect it has great advantage over autonomous mode [5,6].

In conclusion, the wireless Mesh network have both traditional fat AP and thin AP's advantage, which is a convenient network, that can support multi-hop, have high capability and high speed and can conveniently provide robust, reliable support for mobile learning network coverage. Mesh wireless campus network is shown in Figure 1. Mesh backbone network are made up of campus regional nodes.

2.3 Campus Wireless Network Coverage Area Planning

Although data transmission speed in the 802.11n standard and Mesh network efficiency are improved, the correct planning of campus wireless network coverage area and access point location can represent the high performance of mobile learning wireless campus network, which would let the methods such as real-time teaching be fully used.

Indoor Coverage. Indoor coverage area size and structural complexity tends to vary greatly, according to the school's specific needs, design a variety of indoor coverage solutions. Generally speaking, for the large indoor environment, such as a library reading room, international conference hall, gymnasium, large classrooms, which have more centralized network users, to design with a single AP (Access Point, wireless access point) and small area coverage is recommended. Multiple AP integration cross cover forming a wide coverage area, each AP independently connect to the switch fabric, to ensure the effective bandwidth is wider, In the implementation of specific projects, generally we use indoor wireless AP products compatible with the 802.11n standard to complete wireless network indoor coverage [7].

Outdoor Coverage. Mainly use outdoor AP to finish target signal coverage, the school stadium, center square, teaching building public areas, generally are required to achieve the wireless coverage. According to actual situation of the outdoor area, we can design a number of wireless base stations, using overlapping radio to cover the area, in order to complete seamless wireless coverage and provide a good support network for the school to organize large-scale activities.

Comprehensive Coverage. The buildings have both outdoor and indoor scenes in the school, covering wireless signals. For office buildings, teaching buildings and other more complex structure region, according to the specific situation of building structure, there are two choices: the first is using wireless AP products with high sensitivity, strong penetrating ability, and separate type ceiling antenna, to complete the interior regions of the complete coverage by the way of a AP with an antenna, or a

AP with multiple antenna. A dedicated wireless AP, with indoor ceiling antenna, is used to complete the internal wireless coverage of the buildings. Using separated antenna design, can adapt to the use connect the wireless equipment with high gain antenna connection, to ensure the wireless signal with high quality can cover more distance, meanwhile to enhance the ability of equipment in the greater interference environment. The second is Using the outdoor covering methods, select outdoor wireless AP, gather wireless signal by antenna, so that the wireless coverage become larger, penetrating ability become stronger. Equipment and antenna are placed at the top or bottom of the building, to transmit the wireless signal downwardly or upwardly in order to cover the whole building [8].

2.4 The Management of Wireless Campus Network

Radio Frequency Management. Wireless signal in space transmit freely, wireless environment is also changing, considering the interference among channels in wireless signal propagation, we need to do the wireless campus network radio frequency adjustment constantly in order to adapt to the change of wireless users' environment. In the past, wireless campus network radio frequency management relied on engineer's manual configuration. This kind of fixed, static management means is not flexible, has great limitations, especially not suitable for large-scale wireless network deployment and dynamic complex wireless environment, therefore, we need a more intelligent and dynamic frequency management.

After gathered by the wireless switch, unified through intelligent control, RF information between AP and AP can be very convenient to automatically adjust all AP wave characteristics in the line, which can reduce the interference between AP. Automatic configuration can be set in a specific time, such as 2 am, there would be little users.

2. The User Management.

User group classification

Campus wireless user group mainly includes the following group: the teaching, research and management staff; college students, graduate students, doctoral students and international students; short-term training teachers, visiting scholars.

User authority management

Set different user authorities according to the different user groups, and set different access authorities according to the different user authorities.

Unified authentication management

The wireless network and wired network should have unified identity authentication, teachers and students can access campus network through the unified authentication, to achieve wireless roaming between campuses.

As shown in Figure 2, wireless network connect with the existing wired authentication network platform, and due to the use of same user information database, the wireless network and wired network can use the same client software of 802.1x, or Web Protal authentication to access the internet.

2.5 The Security Deployment of Wireless Campus Network

The channels for wireless network data transmission are open. Safety problems must be considered, there are three aspects that we should consider to ensure the safety:

identity authentication, data encryption and data integrity. At present, WAPI and IEEE 802.11i are leading wireless LAN transmission protocol at home and abroad.

WAPI. Wireless LAN Authentication and Privacy Infrastructure(WAPI) is a kind of security protocol, and also China's wireless LAN security mandatory standard. The core content of WAPI is authentication of unique WAI protocol based on WAPI, using a certificate as the identity certificate; Use SMS4 algorithm for data encryption; complete unicast key negotiation based on the 3 handshake process, complete multicast key agreement based on the two handshake process [9].

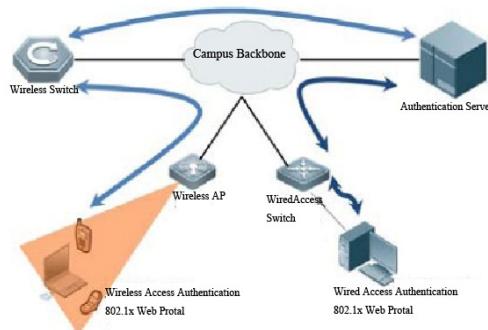


Fig. 2. Unified identity authentication in wireless and wired network

802.11i. The core content of 802.11i includes identity authentication based on 802.1x; encryption based on the TKIP, CCMP encryption algorithm; the achievement for user session key dynamic negotiation based on the 4 handshake [10].

For the identity authentication, 802.11i and WAPI both support digital certificate authentication, can realize the wireless campus network access device and a wireless client bidirectional authentication. For data encryption, WAPI forced the user use digital certificates as identity documents, which is convenient for safety management, and also improves the security. 802.11i supports a variety of encryption methods, adopt the same encryption mode in communication, in access environment of a variety of encryption methods coexist, broadcast / multicast packets can only use the lower level of security encryption for transmission, which can easily make mistakes. For the integrity of the data, WAPI specifically ensure the integrity of authentication information. In 802.11i, differential success information does not contain the integrity verification, authentication information can be revised and distort, which can become the security threats to the network.

2.6 QoS Deployment

The QoS deployment for wireless campus network which can support mobile learning can increase classification and priority mapping based on user state flow. By distinguishing the various parameters of data flow, flexible QoS strategy including such as source / destination IP address, protocol, service, set the stream classification, priority tag, rate limiting, queue scheduling, coordinated with the cable network QoS to realize end to end QoS of the whole network for a variety of business develop at

the same time. Thus more effectively serve for the application of mobile learning service in the school [11].

The global QoS deployment can consider bandwidth management based on each user, bandwidth management based on each SSID and resource weight management based on each SSID, for example, set STP and RTP protocol at a higher queue, and generally put in the lower queue. So the businesses including voice, video and so on, which are sensitive on time delay, will be able to get the best bandwidth and transmission quality assurance.

2.7 IPv6 Support

For wireless campus network which can provide mobile learning platform, the significance of IPv6 is not only to replace IPv4, only enrich IP address resource. Every student, teacher, every classroom, teaching facilities, even every book are distributed IPv6 address through IPv6 protocol, objects become an organic whole, add additional performance and function, to make teaching and learning interaction transfer from pure screen control to Individualization when using the wireless network.

In the concrete construction, because a lot of wireless devices that do not support the IPv6 protocol, we can make the IPv6 protocol packet transmit transparently, after certificated by IPv4 protocol, connect with upper level IPv6 equipment automatically distribute IPv6 address and route, in order to access IPv6 resource.

3 Application Requirements for Campus Mobile Learning

3.1 Terminal Support for Campus Mobile Learning

The campus mobile learning terminal which use wireless technology to access Internet are terminal equipment for mobile learning in campus network, whose main function is mobile learning, so the wireless network is very important. The mobile Internet devices such as intelligent mobile phone, notebook computers, flat-panel computer can be used as mobile learning terminal.

3.2 Systems for Campus Mobile Learning

At present the main mobile intelligent systems are Android, IOS, Windows Phone and Symbian, Whether network environment, portability, generality can meet the majority of teachers and students learning needs or not is the key content we should consider in the development of the application of mobile learning.

3.3 The Platform for Campus Mobile Learning

Most of the universities have established service systems based on the network for office automation, integrated educational management, financial management, et.al., to provide various information services. The combination between present system and

the wireless network platform will become the difficulty of construction. Create an open mobile learning platform, which can support a variety of application protocol and can be managed well, require the system can fully support the typical applications in teaching, scientific research and office and so on. such as mobile phone web browsing, send and receive mail, file transfer, Trans-Campuses bus query, campus map navigation, calendar, course arrangement, and in addition especially multimedia applications which is developing rapidly in recent years, such as real-time or non-real-time VOD, visual conference, instant messaging, IP telephone, teaching blog and interactive communities.

4 Conclusion

With the development of research for mobile learning campus wireless network technology, setting up mobile learning campus wireless network can create mobile learning platform for the teachers and the students to improve their learning ability, research level. In addition, new business applications based on the original digital campus will also appear, so that every teachers and students who has mobile learning terminal, can freely enjoy the application of mobile learning.

References

1. Bingxi, D., Bainian, L., Jinding, M.: Research on Standards and Key Technology of IEEE802.11n in WLAN. *Communication Technology* 8, 123–125 (2008)
2. Hua, W., Jingjing, L., et al.: Research on Wireless Mesh. *Journal of Nanjing University of Information and Engineering (JCRSCI)* 2, 332–336 (2010)
3. Novosád, T., Martinovič, J., Scherer, P., Snášel, V., Šebesta, R., Klement, P.: Mobile Phone Positioning in GSM Networks Based on Information Retrieval Methods and Data Structures. In: Snasel, V., Platouš, J., El-Qawasmeh, E. (eds.) ICDIPC 2011, Part II. CCIS, vol. 189, pp. 349–363. Springer, Heidelberg (2011)
4. Wang, S., Min, J., Yi, B.K.: Location Based Services for Mobiles: Technologies and Standards. In: IEEE International Conference on Communication (ICC), Beijing, China (2008)
5. Yongjian, C., Lin, C.: Wireless Mesh and Its Application. *Network Communication and Security* 1, 3369–3372 (2010)
6. Prommak, C., Wechtaison, C.: On the Quality of Service Optimization for WiMAX Networks Using Multi-hop Relay Stations. In: Snasel, V., Platouš, J., El-Qawasmeh, E. (eds.) ICDIPC 2011, Part II. CCIS, vol. 189, pp. 93–106. Springer, Heidelberg (2011)
7. Xiaobo, W.: Solution of Wireless Campus Network Construction. *China's Information Technology Education* 4, 86–87 (2009)
8. Zhao, Y.: Standardization of Mobile Phone Positioning for 3G Systems. *IEEE Communications Mag.* 40(7), 108–116 (2002)
9. Shujing, W.: Security Research on WLAN Based on IEEE802. 11. *Journal of University of Science and Technology of Hunan* 31(12), 86–88 (2010)
10. Jun, Z., Zhenhai, H., Yuehua, Z.: Communication Performance Analysis on WAPI and IEEE 802.11i. *Security Protocol. Communication Technology* 40(12), 228–231 (2007)
11. Zhida, Z.: Research on QoS Deployment and Application Concerned with Wireless Campus Network. *Guangxi Journal of Light Industry* 12, 85–86 (2009)

Simulation of Malicious Nodes Detection Based on Machine Learning for WSN

Kuancheng Zou¹, Yuanling Ouyang¹, Chuncheng Niu¹, and Yi Zou²

¹ School of Computer Science and Engineering, Changchun University of Technology,
Changchun, Jilin, 130012

² Dalian Fire Protection Engineering Co, LTD, Dalian, Liaoning, 116013
89580483@qq.com, ouyangyuanling@126.com

Abstract. A Machine Learning based malicious nodes detection scheme was designed and simulated in this paper, which has modeled the characteristic behavior of Selective Forwarding, Hello Flooding and abstracted four properties of these two types of malicious nodes closely related. The nodes can be classified by using Machine Learning algorithms. The experimental results show that this method can correctly classify a certain number of sensor nodes with a lower error rate.

Keywords: malicious nodes, Selective Forwarding, HELLO Flooding, Machine Learning, simulation.

1 Introduction

Wireless sensor network (WSN) is composed of a large number of tiny sensor nodes, and forms a multi-hop ad hoc network as the wireless communication mode. Sensor nodes are generally deployed in harsh environment or human could not easily touch them. The nodes can work in the open air environment, and the sensor network topological structure may change because of bad physical conditions or self energy deficiency, and what is more, due to their node self fault, malicious nodes by man-made destruction affect the wireless sensor network communication security, thereby affect the whole wireless sensor network communication quality. The attacker has two kinds of models, one is the mote-class Attackers and the other one is laptop-class. The first kind has a limited ability to influence others, just affecting its neighbor nodes. Compared with legal nodes the second kind of malicious nodes have more advantages in stronger power supply, CPU, high energy transmit power and more sensitive antenna , so that they can affect the whole wireless sensor network. Attack mode has two kinds, which is divided into external aggression and internal attack, so far the attacks mentioned have been external attack. But the internal attack also should be considered for some of the original valid nodes change into the malicious nodes or run malicious codes or steal key code and data by an attacker from legitimate nodes, and then they use the more powerful device attacks in wireless sensor networks. Sensor network attack can be divided into the following categories: Deception (choice),

Selective Forwarding attack, Slot Holes, Sybil attacks, Hello Flooding and confirmation of deception.

The attacks of malicious nodes can be divided into positive and passive patterns. Passive attack is that malicious nodes steal the data of wireless sensor network through monitoring means, this attack can be prevented by data encryption, authentication and other prevention methods. Active attack refers to the malicious nodes via packet dropping let down network communication and through changing routing information affects the normal network operation. The typical patterns of attacks include Selective Forwarding attack, Slot Holes attack, Hello Flooding and others. Malicious nodes in wireless sensor networks will weaken performance, therefore the malicious nodes detection research for wireless sensor network provide the necessary protection to the normal operation. The malicious nodes detection method based on Machine Learning has become a research hotspot.

2 The Research Status

Both Domestic and foreign scholars have put forward various algorithms to detect malicious nodes. Article [2] proposed a malicious nodes detection based on the source code and multi-path transmission. Literature [3] proposed a method of detecting malicious nodes based on classification method through Bayesian classification of samples, to determine the type of the nodes. Literature [4] design a detection scheme based on the distributed of Machine Learning, using K nearest neighbor algorithm to get clusters in the sensor network nodes. Use the Bayesian classification algorithm to anomaly detection in a cluster Article [5] is proposed black-hole and selective forwarding attack detection based on support vector machine SVM and sliding window. Literature [6] is proposed a detection scheme to capture the network intrusion based on a hidden Marco model. Literature [7] proposes the use of a fixed bandwidth clustering algorithm to detect the abnormal node in the intrusion detection system. Literature [9] introduces a intrusion detection model through electing monitoring communication node independent which based on the spontaneous watchdogs. Literature [10] defines a intrusion detection system for selective forwarding and slot hole attack in a sensor network. Literature[11] is introduced a anomaly detection based on security scheme which employs a sliding window to detect flooding and resource exhaustion attacks. Literature [12] puts forward to a kind of selective forwarding attack detection scheme with multi-hop cognitive in base station and the source node. Literature [13] proposed a intrusion detection technology by using minimal cut set and minimum dominating set to realize this algorithm. Literature [14] proposed a simple intrusion detection model which plays a role in the link layer and network layer, via the collision ratio, power, the integrity of the data packet and some other properties, to achieve intrusion detection.

These above documents have provided important reference for malicious nodes detection. But some of them only base on theoretical without actual testing or did not consider the communication between nodes. These documents do not combine Selective Forwarding with Hello Flooding when realized detection research, part of

literatures are on Selective Forwarding attack, slot hole, the wormhole attack in-depth, while the other parts are on Hello Flooding malicious nodes attacks. It usually causes the packet lost when hole attack happened, the wormhole attack, to a certain extent they are on behalf of Selective Forwarding, therefore this article selects Selective Forwarding and Hello Flooding malicious node detection as the research point.

3 Wireless Sensor Network Attack Modeling

For Selective Forwarding, Hello Flooding attacks, the successful transmission rate[3], forwarding rate[3], the average received power, average transmission delay are the related attributes of wireless sensor node attribute modeling. Because the wormhole attack or slot hole attack usually accompanied with Selective Forwarding, therefore this article only considering Selective Forwarding and Hello Flooding for the sake of simplicity. As everyone knows, Selective Forwarding is that malicious nodes according to a certain proportion of discarded or forwarding data packets. For Hello Flooding, malicious nodes are to use of its powerful transmitter power to send hello packets, let the network within the region of the nodes are mistaken for the malicious nodes are their neighbor, then send packets to the malicious nodes. The average received power and average transmission delay rate can be used as Hello Flooding test attributes.

Successful Transmission Rate:1- Packet loss rate[16].

Forwarding Rate: If a node receives ri packets, has forwarded fi packets, then this node's forwarding rate is $\frac{ri}{fi}$.

The Average Received Power: If there are N_i neighbors receiving the packet when a node broadcasting, this node's average received power is $\frac{N_i}{\max(N_1, N_2, \dots, N_{500})}$.

Average Transmission Delay Rate:If a node sends a data at the average delay T_i , then the node's average transmission delay rate is $\frac{T_i}{\max(T_1, T_2, \dots, T_{500})}$.

4 Detection Scheme Based on Machine Learning

Bayesian classification are one of the ten effective Machine Learning algorithm, which is a kind of unsupervised learning, with prediction accuracy, speed, computational convenience features and so on [4]. Bayesian classification algorithm's principle is through the prior probability of an object, Bayes formula to calculate posterior probabilities, i.e. Which a certain class is the object belonged to depends on the maximum probility.

If there are n classes {C1,C2,...Cn},to the object X,if known $P(C|X)$, X is belongs to the class C,for any $i \neq j$,if $P(C = C_i | X) > P(C = C_j | X)$,then X belongs to class Ci. We use Bayes formula

$$P(C|X) = \frac{P(C) \cdot P(X|C)}{P(X)} \quad (1)$$

$P(C)$ is a priori probability, $P(X|C)$ is the Similar Function, $P(C|X)$ is the posterior probability density.

It should obtain a priori probability density first as for the known types of samples $\hat{P}(Ci)$ and Similar Function $\hat{P}(X|Ci)$,then use discriminant function of the posterior probability density: $gi(x) = \ln \hat{P}(Ci|X)$.

For any unknown sample X, using the above functions, according to the maximum posterior probability as the sample X's classification $P(Ci|X) = \max p(C_k|X)$, $k \in \{1, 2, \dots, n\}$,then $X \in Ci$.

5 Simulation

Using NS2 as a simulation experiment platform, version 2.35.MAC protocol is ZigBee (IEEE 802.15.4), routing protocol is AODV, the whole simulation last 50s, 0.5s maximum pause time, maximum speed is 20m / s, the rate is 1.0 and the two types of malicious nodes including Selective Forwarding and Hello Flood attack. Specific test parameters shown in table 1 as following:

Table 1. Simulation parameters

Paramet	value
Size	800x800
Number	500
Type	cbr
Seed	1
MC	100
Data rate	1.0
Pause time	0.5
Routing	AODV
MAC	IEEE 802.15.4

According to the specific test settings, nodes in the scene random distributed shown in Fig. 1 as following:

It has known 50 sensor nodes's type, which contains 2 malicious nodes. In the rest there contain 3 malicious nodes. Nodes can not move here and there in the scene space.

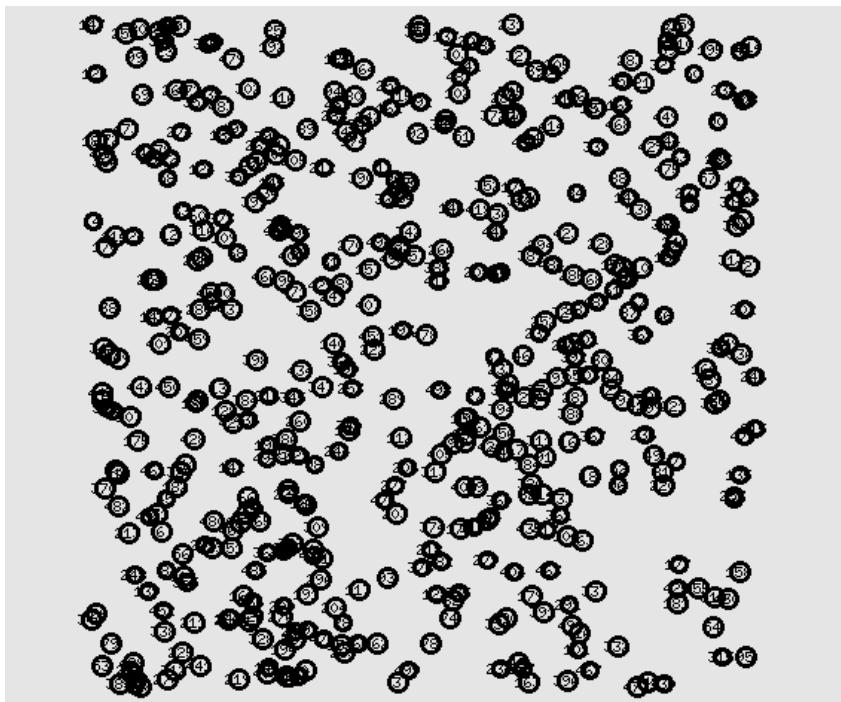


Fig. 1. Sensor nodes random distribution map

5.1 Program Code

Writting a TCL file, includes network topology and data transceiver process and so on.

```
# Definition of scene parameters:
.... ...
Set val(netif)          Phy/WirelessPhy/802_15_4      ;
set val(mac)            Mac/802_15_4                   ;
  set val(ifq)          Queue/DropTail/PriQueue     ;
set val(ll)             LL                           ;
.... ...
set val(nn)              500                         ;
# number of mobilenodes
set val(rp)              AODV                        ;
# routing protocol
  set val(x)            800                         ;
set val(y)              800                         ;
set val(cp)              "cbr5"                      ;
set val(sp)              "set5"                      ;
```

Then create 500 sensor nodes, including 5 malicious nodes have defined:

```
for {set i 0} {$i < $val(nn)} {incr i} {
    set node_($i) [$ns_ node $i ]
}
.... ....
$ns_ at 0.0 "[${node_(5)} set ragent_] abc"
$ns_ at 0.0 "[${node_(25)} set ragent_] Fooled"
$ns_ at 0.0 "[${node_(105)} set ragent_] abc"
$ns_ at 0.0 "[${node_(275)} set ragent_] abc"
$ns_ at 0.0 "[${node_(445)} set ragent_] abc"
.... ....
```

In these above codes, string "abc" respents Selective Forwarding, while Hello Fooled is on behalf of Hello Flooding. Finally run the TCL file and produce the map of sensor nodes random distribution as the above Fig.1 .Using gawk command abstract attributes from the data file, then produce a sample set. For the sample set using the Machine Learning algorithms previously described to learn, generate discriminant, classify the remaining unknown type of nodes.

6 The Simulation Results and Analysis

The nodes are classified by using Machine Learning, the results obtained as shown in fig. 2:

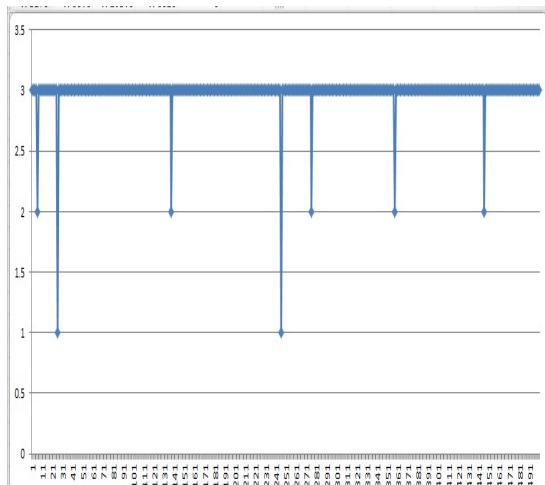


Fig. 2. The map of classifying nodes type

The graph X axis is nodes number, Y axis is nodes type, number 3 represents the normal nodes, number 2 represents the selective forwarding malicious nodes, number 1 represents Hello Flooding malicious nodes. From above, it can draw the test detection rate and false positive rate.

7 Conclusion

According to the Selective Forwarding and Hello Flooding feature attributes extraction for four kinds of properties of these two kinds of malicious nodes detection in this paper. Compared with other literatures, this paper has more advantages in the number of simulated sensor nodes, in which the attributes values extraction simultaneously two kinds of malicious nodes in wireless sensor network communication behavior. The experimental results show that this method can correctly classify a certain number of sensor nodes at a 0.6% error rate.

References

1. Karlof, C., Wagner, D.: Secure Routing in Wireless Sensor Networks: Attacks and Countermeasures. In: 2003 IEEE International Workshop on Attacks and Countermeasures, pp. 113–127. IEEE Press, New York (2003)
2. Xu, J.-H.: Wireless sensor network malicious nodes detection and localization strategy research. Central South University, Hunan (2011)
3. Liu, H., Cui, J., Dai, H.: Malicious nodes detecting in wireless sensor networks based on multivariate classification algorithm. Journal of Sensor Technology 24, 771–777 (2011)
4. Xiao, Z., Chen, Z., Li, Q.: anomaly detection simulation study based on distributed machine learning for WSN. Journal of System Simulation 23, 181–187 (2011)
5. Kaplantzis, S., Shilton, A., Mani, N., et al.: Detecting Selective Forwarding Attacks in Wireless Sensor Networks using Support Vector Machines. In: ISSNIP, vol. 10, pp. 335–340 (2007)
6. Yang, X., Wang, C., Wang, D.: Intrusion detection system based on HMM. Computer Engineering and Applications 12, 149–151 (2005)
7. Loo, C.E., Ng, M.Y., Leckie, C., et al.: Intrusion detection for routing attacks in sensor networks. International Journal of Distributed Sensor Networks 2(4), 313–332 (2006)
8. Perkins, C., Royer, E.: Ad-hoc on-demand distance vector routing. In: WMCSA 1999: Proceeding of the Second IEEE Workshop on Mobile Computer Systems and Applications, p. 90. IEEE, NJ (1999)
9. Roman, R., Zhou, J., Lopez, J.: Applying intrusion detection systems to wireless sensor networks. In: CCNC 2006: Proceeding of the 3rd IEEE Consumer Communications and Networking Conference, pp. 640–644. IEEE, NJ (2006)
10. Krontiris, I., Dimitriou, T., Freiling, F.C.: Towards intrusion detection in wireless sensor networks. In: EW 2007: Proceeding of the 13th European Wireless ConferenceL Enabling Technologies for Wireless Multimedia Communications, ENSTA, Paris (2007)
11. Onat, I., Miri, A.: An intrusion detection system for wireless sensor networks. In: WiMob 2005: IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pp. 253–259. IEEE, New York (2005)

12. Yu, B., Xiao, B.: Detecting selective forwarding attacks in wireless sensor networks. In: IPDPS 2006: 20th International Parallel and Distributed Processing Symposium, IPDPS, Greece, pp. 8–15 (2006)
13. Anjum, F., Subhadrab, D., Hu, et al.: On optimal placement of intrusion detection modules in sensor networks. In: BROADNETS 2004: Proceedings of the First International Conference on Broadband Networks, pp. 690–699. IEEE, San Jose (2004)
14. Xiao, D., Chen, C., Chen, G.: Intrusion detection based security architecture for wireless sensor networks. In: ISCIT 2005: Proceeding of the International Symposium on Communications and Information Technologies, pp. 1365–1368. IEEE, WuHan (2005)
15. Routing Security in Sensor Network: HELLO Flood Attack and Defense, <http://ftp.tcc.edu.tw/docs/Internet-Drafts/draft-suhopark-hello-wsn-00.txt>
16. Ke, Z., Cheng, R., Deng, D.: NS2 simulation experiment multimedia and wireless network communication, pp. 111–123. Publishing House of electronics industry, Beijing (2009)

A Robotic System Based on Wireless Control

Changjiu Pu

Network Management Center, Chongqing Education College, 400067 Chongqing, China
pcj8880289@sina.com

Abstract. In order to show and gradually control, an embedded robotic experimental system based on frequency modulation in the paper is proposed, which consists of remote control and robot body. The systematic hardware structure and realization principle is introduced in detail in the paper: master MCU of remote control send control information to slave MCU of robot body, slave MCU returns the message after it controls the motors to realize, and master MCU send next control information again after accepting the message. The system saves the controlling code of the robot in the remote control and achieves separation of code and robot. The effectiveness of the system is validated by the theory analysis and the experimental results.

Keywords: frequency modulation, embedded system, robot, control method.

1 Introduction

Robot is an information technology, automatic control technology, sensor technology, robotics and electronic technology in a scientific and technological achievement. Robotics is a high degree of cross-frontier discipline, which causes great interest in people with different professional backgrounds such as mechanical science, biology, anthropology, computer science and engineering, control theory and control engineering, electronic engineering, artificial intelligence and sociology. Therefore, Robot as a research object is high academic value to building teaching platform and meets the need of Complex and innovative talent [1, 2].

Now, the robot is mainly driven by motors. Especially, the biped robot with higher flexibility and mobility need more motors to control [3, 4]. But the communication of biped robots is an important component of the control system of the robot. The way with communication cable is limited. If instead of using infrared communication way, the directional of infrared transmission will affect the movement of the robot in a certain extent, which can not guarantee all the data is fully received while walking and affect the robot control. The remote control technology of wireless frequency modulation needn't equip the special-purpose communication cable and basically have not the question of directionality. Especially the appearances of the wireless frequency modulation module, it reduced interference, simplified the design of wireless communication of robot, increasing the dependability of communication, making the system simple, cheaply and so on.

Therefore, the paper studies the biped robotic experimental system based on wireless frequency modulation for hardware, which modularizes control procedure of the robot in controller, need not use computer in communication process, and increases the dependability and easy to show and gradually control.

2 Hardware Design of Robot

This system is mainly divided into two major parts: robot body and the remote controller. Height of the robot is 260mm, weight is about 1240g. It consists of truck, arm, leg and sole, equipping with controls chip and battery. Robot has 17 degree of freedom, waist has 1 degree of freedom, each arm has 3 degree of freedoms respectively and each leg is 5 degree of freedoms. Robot body and every electrical machinery serial number are as shown in Fig. 1, design of bigger sole, improving carrying capacity and flexibility of the robot and conducive to simplify balance control of the robot.

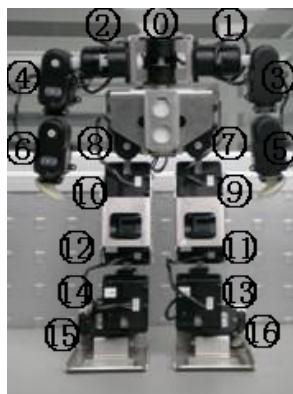


Fig. 1. Picture of robot body

2.1 Robot Body

The whole system adopt master-slave control structure, control part of which is make up of consists of one master MCU(microprocessor control unit) and 17 slave MCU. The master MCU is the Hitachi 50 MHZ and 32-bit microprocessor (SH2-7145). The 32-bit RISC microcontroller for the microprocessor, the chip is Harvard architecture for separating from program and data, built-in I2C bus, 10-bit A/D, DAM controller, four full-function serial port, 256K ROM and 8K RAM. In the paper, 144-pin LQFP package is used and RAM is expanded to 64M. Robot body supply with +9.6 V lithium battery, which makes the robot work full effect.

Because SH2-7145 has strong data processing capabilities, it is the control center of whole system. It runs the main program, receives the control order from serial port from remote controller, finishes relevant data processing, produce speed and angle of each joint every motion and send the speed and the angle to each slave MCU by system bus for controlling motors and operation of the system. Slave MCU and DC

motor is match, which receive the control information (such as angle and speed) from master MCU, and operate and deal with the control information.

Because the servo motor of direct current has good linear characteristics [5], the start torque is relatively larger; it is wide to adjust speed range, high efficiency and great control the characteristic [6]. Meanwhile, the control circuit of the servo motor is relatively simple, thus reducing volume and quality of the servo controller, promoting the integration of the control circuit and robot; And the control algorithm of the servo motor of direct current of permanent magnetism is simple, which promote to real-time in coordination of the robot control system. So, AI-701 was chosen.

The motor changes the speed information into control current of DC by the circuit of PWM (pulse width modulation), controls the rotational speed of the motor by size of the electric current, detects and gets feedback information by the built-in sensors, compare with signal which the master MCU send. All of that make the motor turn into the correct position.

2.2 Design of the Remote Controller

Power supply of control system and power supply of motors are separate. Control system (the remote controller) adopts 80C51F330 which is an upgrading edition of 51 MCU. 80C51F330 adopt a pair of wireless frequency modulation module (433M) to realize sending and acceptance of wireless signal. In the part of robot body, the voltage (9V) that the battery offers Supplied the module of frequency modulation for +5V by 7815 steady voltage, and then +5V lowers to +3V and supplies MCU by 1N5226 steady voltage diode. In the part of the remote controller, keyboard adopt ranks scan way. In the sender of the system, signals which its need is required to send 5th time in succession. In the accepter of the system, it is correct to accept until two same signals continuously accepted. Acceptor of the system sends the signal to the robot by serial port, the baud rate of serial port is 2400. The main parts of circuit diagram are as shown in Fig. 2.

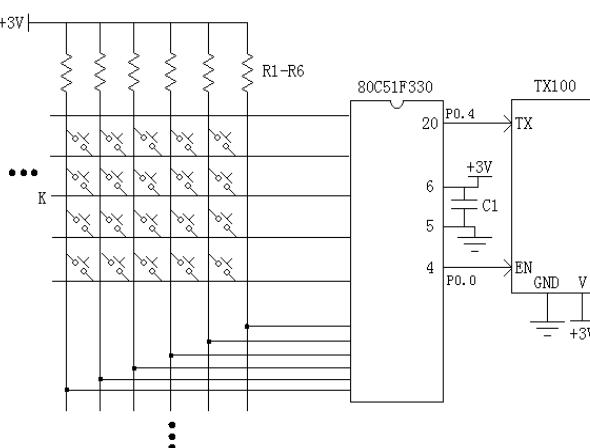


Fig. 2. The main parts of circuit diagram of the remote controller

3 Design of Software

Robot control algorithm takes product inference engine as fuzzy control [7]. Because of cost considerations, the robot does not use sensor such as accelerometer, gyroscope. So it can not take the current state of the robot as input. Therefore, manual input will be used.

First, a few basic movements of the robot will be plan [8]. Then, parameters (angle, speed, etc.) of the 17 motors of robot will be calculate from the own characteristics of multi-degree of freedom control of robot and control theory based on a mathematical model. Parameters of the 17 motors of robot need be inputted when programming. The robot will take this data to calculate by fuzzy inference engine [9] and obtain various angles of motors which allow smooth implementation of motion planning of the robot and do its best to meet the requirements of the various motor angles. The angle will be used to control the motors to achieve the stable motion planning.

Principle of motion design of the robot is: First, consider the feasibility of planning designed to ensure that planning can be successful (The robot does not fall); Second, consider the coordination of planning to make the planning look close to people's action; Last, consider the limits of the motor, which is the motor can turn the maximum angle.

Master MCU send control information (angle, speed, etc.) to slave MCU, slave MCU returns message after it controls the motors to realize the planning, and master MCU send next control information again after accepting the message.

The main program process is as follows:

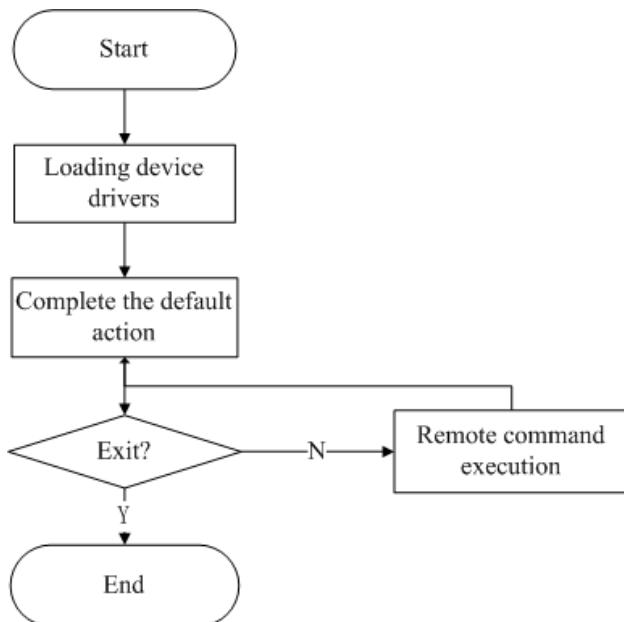


Fig. 3. Flow chart of main program

The following are block diagram of program for sending and receiving:

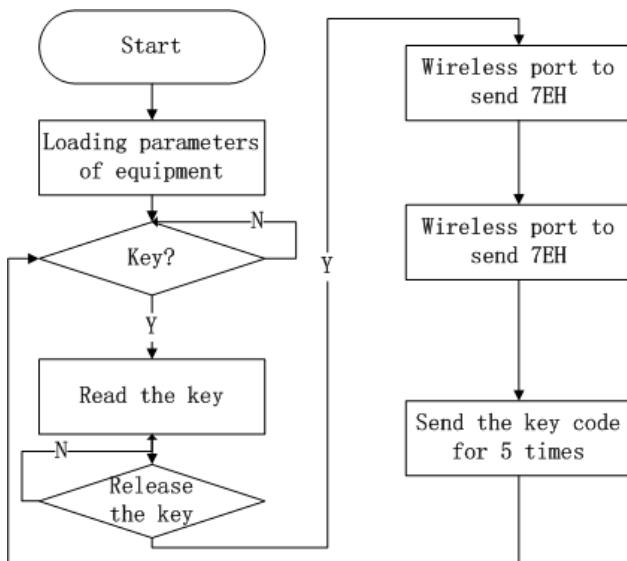


Fig. 4. Flow chart of program for sending and receiving

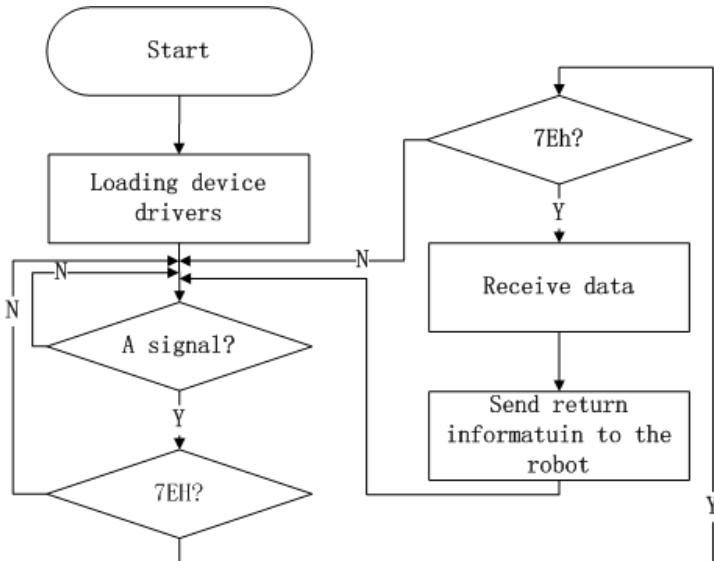


Fig. 5. Flow chart of program for sending and receiving

From fig. 4 and fig. 5, 7EH takes as lead code, which can reduce misuse because of pressing down the control button accidentally and errors are caused by hardware problems.

Control program compiled with the C language. The Programming is very simple. A subroutine for opening the legs is as follows:

```
void leg_open (int ko, int tim, int op)
{
    op = op / 10;
    west_down (ko, tim);
    wr_xz (0, 5 + ko);
    txt = 1; txx8 = txx = 0;
    while (txx <= op) l_open (-txx * 100);
    wr_xz (0, ko);
    st_legs ();
    west_up (ko, tim);
}
```

4 Experiment

In order to verify the validity of the system in the paper, the planning motion in the paper [10] was used in the robot. Using off-line programming method, according to the basic posture of robot in the planning, proper restrict center of gravity trajectory, to ensure equilibrium conditions on ZMP and proper geometric constraints the joint angle information (degrees of freedom needed) can be solved [11].

In the planning, angle of each joint is set to the initial angle (Value of the angle of each joint is 0 degree) when the robot stand straight. Then, size of each joint each state are be calculated accord to planning. Angle information is converted and written into the remote controller.

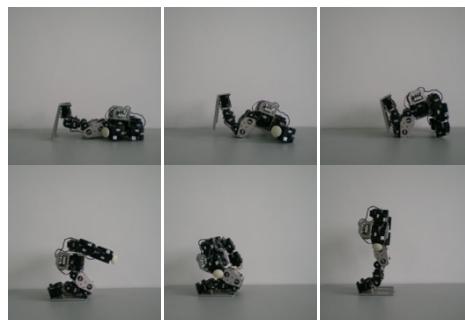


Fig. 6. State diagram of experimental

5 Conclusion

This paper presents an embedded the robotic experimental system, which consists of remote controller and receiver based on frequency modulation to control 17 joint

motors of robot. Implementation principle and implementation method to achieve is introduced in detail. The robot can not only implement the planning which is designed in advance, but also implement the planning action in accordance with operation in the wireless remote controller. It realizes the separation of robot code and robot. The system overcomes the shortcomings in cable technology and infrared technology. This system need not to use the computer, simplify the design of wireless communication of robot and increase the dependability of communication. Meanwhile, it is cheap, advantage of showing and gradually controlling. Theory analysis and experimental result have verified validity of the system.

Acknowledgments. The work described in this paper was supported by a Grant from Chongqing Education College (No.KY201175C).

References

1. Chen, X.-P., Luo, W.-J.: A study of the practicum and Innovation training system Based on Robot Experiments. *Education and Modernization* (3) (2011)
2. Wang, J., Ma, H.-X., Huang, Q.-W., Liu, J.-P.: Study on Robotics Teaching for Undergraduates Education. *Journal of Electrical & Electronic Education* 33(2) (2011)
3. Bao, Z., Ma, P., Jiang, S., Cheng, J., Wang, C.: The problem and the research history from biped robot to humanoid robot. *Robot* 7 (1999)
4. Xie, T., Xu, J., Zhang, Y., Qiang, W.: History, current state, and prospect of study of humanoids. *Robot* 7 (1999)
5. Xu, W.-X., Ruan, R.-L.: Design of Wireless Digital Control System for DC Servo Motor. *Journal of Shenyang Agricultural University* 41(1) (2010)
6. Tan, J.: Dedicated motor control circuit. *Machinery Industry Press* (August 1997)
7. Wang, L.: Fuzzy systems and fuzzy control tutorial. *Tsinghua University Press* (2003)
8. Chen, Y.: Research on Motion planning and stability control method of Humanoid Robot. *National University of Defense Technology* (2010)
9. Liu, K., Zhang, F.-J.: The hierarchical fuzzy control of 2-DOF robot. *Manufacturing Automation* 33(20) (2011)
10. Fu, G., Yang, Y., Li, J.: Survey and Prospect on Walking Control Strategies for Humanoid Robot. *Machine Tool & Hydraulics* 39(23) (2011)
11. Pu, C., Wang, Y., Li, Z., Liu, W., Qiu, Y.: A method for trajectory planning in the stability of getting up for biped walking robots. *Southwest University (Natural Science)* 30(11) (November 2008)

Study on Wireless Campus Mesh Routing Protocol Based on Security Association

Limei Fan

Department of Information Engineering, Jilin Business and Technology College 130062,
Changchun, China
f1mf1m123@163.com

Abstract. In order to enhance the security of security routing protocol, this paper combined with the wireless Mesh network campus its own characteristics, design a kind of security association wireless campus Mesh network routing protocol -SAWCMRP. The protocol is the use of IKE key exchange technology in the source node and the destination node to establish shared security association SA, This can ensure the sending and receiving message integrity, availability and undeniable. Introduced the SAWCMRP algorithm design process, Through examples validation, SAWCMRP protocol can guarantee the intermediate node cannot be tampered with, counterfeit message, thereby avoiding the worm, hole malicious attacks, at the same time and it doesn't affect the performance of the network.

Keywords: Security Association, Wireless Mesh Network, Route Protocol, IKE, SAWCMRP.

1 Introduction

With the rapid development of information technology, Internet provides Internet and open make information exchange and sharing become reality, However, the information exchange is controllable, sharing is authorized, therefore the information security and reliability issues have become increasingly prominent, and has attracted the attention of the whole world. First of all, the wireless network than the wired network more vulnerable to attack, as everyone knows wireless communication are vulnerable to eavesdropping, information tampering and denial of service attacks. Secondly, the wireless Mesh technology intrinsic multi-hop, self-organized nature, further magnified in wireless exist network security risks, How to find out an effective and feasible safety routing solution is an important topic for wireless Mesh network safety domain , this paper on the security of routing protocols for study.

2 The Wireless Mesh Network

The wireless Mesh network (WMN) is one kind of more than jumps, has the organization and the self-recovery characteristic wide band wireless network

architecture, namely one kind of high-capacity, high speed distribution network [1], It is different with traditional any kind wired and the wireless network, may regard as is WLAN (jumps only) and moves Ad the hoe network (jumps much) the fusion, and has displayed both superiority.

What the Mesh network main is one kind of network construction thought that the major function manifests is not having the center, from the network, to jump much the connection, the route judgment choice and the stable property and so on. The wireless Mesh network is one kind may provide the omnipresent wide band service the low power loss, the inexpensive solution. The WMN structure mainly has three kind of [2], namely the wireless backbone Mesh network architecture (Backbone WMN), has Mesh the Clients terminal wireless backbone Mesh network architecture (Clients WMN) and based on the Hybrid wireless Mesh network architecture (Hybrid WMN).

3 Wireless Mesh Network Route Protocol

Ad hoc on-demand Distance Vector, AODV [4], is a reciprocal, based on the purpose of on-demand route protocols. It is essentially dynamic source route protocol DSR protocol and DSDV comprehensive, it borrows the DSR route discovery and route maintenance in the foundation program, the purpose of using the serial number of DSDV in technology, combined with DSR in the thinking and improved demand route. What Is different with DSDV, it and does not maintain a route list, but in needs time only then starts the route discovery process, therefore reduced the route protocol expenses greatly.

The AODV route protocol mainly includes the route discover and the route maintains two processes, as shown in figure 1.

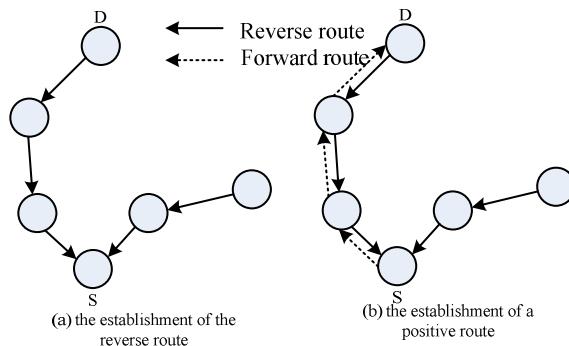


Fig. 1. AODV route builds process

AODV the algorithm belongs to on demand route algorithm, the node does not need real-time to maintain the entire network the analysis situs information, only works as the source node to need to the goal node transmission data packet, only will then initiate the route setup procedure, the route information has not been preserved, has certain life cycle.

AODV protocol band width use factor is high, can make the response promptly to the network topology change, and simultaneously has also avoided the route ring circuit phenomenon occurrence.

4 IKE Protocol

Internet key exchange [5] (IKE) protocol is a consistent with the Internet Protocol Security (IPSec) standard protocol, used dynamic negotiation and establishing security association SA, so as to solve the in an unsafe network environment safety to establish or update shared key problems. The IKE protocol uses Internet Security Association and key management protocol (ISAKMP) definition the basic framework and the OAKLEY protocol of key exchange mode and SKEME protocol of sharing and key update technology, to define their identity authentication and key exchange technology.

However, IKE protocol in the negotiation and establishment of communication both the security channel a very important step is the mutual authentication of the identity of each other. The IKE agreement does not specify a specific identity authentication technology, but through the " proposal load" to support different identity authentication method, used a total of 13kinds of load to complete the consultation and management of message, Part main load specific function as is shown in table 1:

Table 1. The concrete function of partial loads

Load name	Load value	Load description
SA load	1	Consultation security attributes, indicates the interpretation domain DOI domain and the current negotiations situation
Key exchange load	4	Support variable key exchange techniques, such as Oakley, DH, RSA etc.
Identity load	5	In exchange for identity information specific to DOI data, determine the identity of Communications
Certificate load	6	Provides a method through ISAKMP transfer certificate and relevant information
Hash load	8	Hash function to generate data, can verify the integrity of a message
Signature load	9	Digital signature function to generate data, provide undeniable service
Nonce load	10	A provide effectiveness and avoid replay attack big random numbers

5 The Description of SAWCMRP Protocol

SAWCMRP is in the basis of AODV_D routing protocol, through design between source node and the target node safety association SA, to protect the safety of the sending and receiving data packets, and no impact on the overall performance. In this way, route request packet RREQ can verify arrive at the destination node, and a route reply packet RREP can be a source routing completely returned to the source node. But, by the AODV_D routing protocol process knowable, each terminal client node in access wireless campus Mesh network, are to exchange data and access server, Assume that IKE agreement during the process of execution, so that each client terminal nodes are associated with the access server are established corresponding to the shared SA. And between the nodes of A kind of (N jump range) is through the sharing A kind of nodes forward AREQ packet and update your drive watch at the establishment, and save the drive list. This B client terminal node has only one shared SA between the access servers, while A client terminal node has multiple shared SA. If the source node S and the destination node D establish a shared key, To express with K_{sd} .

The SAWCMRP protocol is directly on the AODV_D routing protocol extended, with IP options in the underlying protocol head add an SAWCMRP head, as shown in figure 2. The SAWCMRP protocol also designed query identifier, used to protect the query message and prevent possible pairs of the DoS attack of terminal nodes. Finally, A kind of node in the forward and processing the request for message routing neighbor node has priority mechanism, efficiency improved the stability of SAWCMRP protocol.

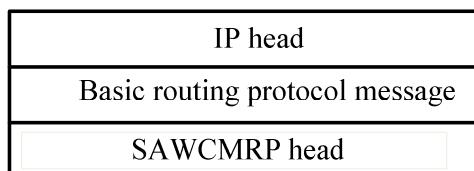


Fig. 2. The message format of SAWCMRP

6 Route Request Packet Structure

6.1 Packet Structure

Terminal client node establishes each shared SA all initiates a query sequence number (Query sequence number Q_{seq}), Q_{seq} as source node S generated routing request increased, when increased to maximum need to update the SA. For each egress route request packet RREQ, the source node S safety pseudo random number generator to generate a pseudo-random query identifier (Q_{id}), the intermediate nodes according to

the Q_{id} to distinguish the different request message. Q_{id} is generated by one-way function, the attacker in limited time is difficult to calculate out, also can't predict, this can effectively prevent the attacker forged request packet. Q_{seq} , Q_{id} and news authentication codes MAC with on SAWCMRP head, its format as shown in table 2.

Table 2. The head format of SAWCMRP

type	Retain
The query sequence number Qseq	
Query identifier Qid	
SAWCMRP message authentication code MAC	

The Type domain is used to identify the different types of SAWCMRP message, such as an access request message AREQ, The route request message RREQ, route reply message RREP, routing error message RERR. MAC is generated by encrypted Hash algorithm [6], the input is message not variable part, including the source node IP, the destination node IP, query sequence number Q_{seq} , query identifier Q_{id} and shared key K_{SD} . While the variable portion can be input,, such as hop, routing node list, otherwise unable to verify MAC.

$$MAC_RREQ = \text{Hash}(IP_S, IP_D, Q_{seq}, Q_{id}, K_{SD}) \quad (1)$$

6.2 The Established Process of Routing

The destination node D receives route request packet, first verify that the RREQ message if there is a security association; then, comparison of Q_{seq} and S_{max} , S_{max} is the source node S in SA within life the existing maximum query sequence number. If $S_{max} \geq Q_{seq}$ may be expired packets or retransmission packet, discard the request packet. Otherwise, the destination node using Hash function count of MAC, and comparison the value with SAWCMRP head MAC, it can verify the RREQ message integrity. After verification pass, node D structure route reply packet RREP. The SAWCMRP head filled with Q_{seq} and Q_{id} , the source node S use these two domains to verify the RREP fresh activity; and then calculating the response message MAC_RREP and fill to the SAWCMRP head the corresponding domain.

$$MAC_RREP = \text{Hash}(Q_{seq}, Q_{id}, List, K_{SD}) \quad (2)$$

But the source node S received RREP packet, first check the source address and destination address, and Q_{seq} and Q_{id} ; Then use the response packet Q_{seq} , Q_{id} , K_{SD} and routing node list calculate MAC, and comparison with packet of MAC_RREP. Through verification, between S and D the secure communication channel is established.

7 Security Analysis of SAWCMRP Protocol

The following is a simple wireless campus Mesh network , the 3 Mesh router nodes and 8 user terminal nodes, the M_1 and M_2 node are two malicious node, as shown in figure 3.

When the source node S has data packet sent to the destination node D, run SAWCMRP protocol to establish a communication route, We use the $\{QST; n_1, n_2, \dots, n_k\}$ represents route request packet, Wherein QST represents route request packet SAWCMRP head, $n_1 = ID_s$, $n_k = ID_d$, $n_i (i \neq 1, k)$ is intermediate nodes IP address. Similarly, the route reply with $\{RST; n_1, n_2, \dots, n_k\}$ to denote. The following specific wireless Mesh network attack type [7] to the analysis of SAWCMRP protocol security.

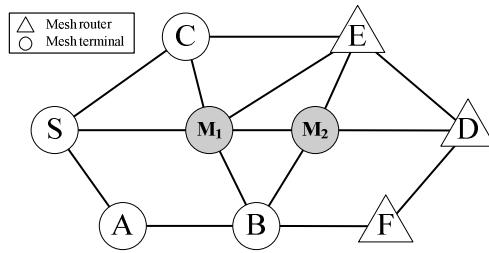


Fig. 3. Simple wireless campus mesh networks topology

(1) Tampering attack

The traditional wireless routing protocol Mesh were assumed to the nodes in the network are mutual cooperation, message forwarding intermediate node does not modify the irrelevant to the routing information, such as the AODV_D protocol. This makes the attacker can easily change the routing request packet or route reply packet any field, thereby generating the wrong route, such as the redirection, loop, resulting in the decline of the network performance. An attacker can tamper with the routing message; the key node is not to route message integrity detection, while the SAWCMRP protocol is a good solution to the problem.

When malicious node M_1 receives route request packet $\{QST; S, C\}$, it altered the request packet for $\{QST; S, X, C, M_1\}$, and forwarded to the neighbor node, Where X is the presence of a node IP address. The request packet to the destination node D, D generates a route reply packet $\{RST; S, X, C, M_1, E, D\}$, and along the routed $\{D, E, M_1, C, X, S\}$ path back to S. When C receives the response packet, it cannot transmit, because of its neighbor nodes does not exist this node X, so the response packet is discard. If the malicious node M_1 correctly routing request packet $\{QST; S, C, M_1\}$, then, finally arrive at the destination node D route request packet is $\{QST; S, C, M_1, E, D\}$. D in reverse path generates a route reply packet $\{RST; S, C, M_1, E, D\}$. When the M_1 receives the RREP packet, it altered the response packet $\{RST; S, C, M_1, D\}$. S received the response packet will be immediately discarded, because it cannot pass through the MAC integrity verification.

(2) Forgery attack

An attacker can forge and broadcast false routing information, such as forged routing request or broadcast an existing routing has been interrupted, it can be caused isolated nodes, the segmentation network, and SAWCMRP protocol can effectively avoid forgery attack.

If malicious node M₁ to the source node S were a period of observation, it forged a route request, the purpose is to let the intermediate node stores the RREQ packet identifier, discarded after S sending legal route request packet. The no security mechanism in wireless Mesh routing protocol, the query identifier is monotonically increasing, is easy to be attacked by. But the SAWCMRP query protocol identifier is generated randomly, in a short period of time is not calculated.

(3) The worm attack

Two or more nodes colluding with the attacker, using tunnel way directly connected, across the normal topology, forwarded directly the routing request and response messages, causing the errors routing topology information.

Malicious node M₁ receives route request packet, with the tunnel way forward to the malicious node M₂, after M₂ received the forged M₁ between M₂ and routing, the routing request message to broadcast , M₂ radio{QST; S, M₁, X, M₂}. The destination node D received in accordance with the route {D, M₂, X, M₁, S} send routing response message. After M₂ received in the form of a tunnel is forwarded to the M₁, finally M₁ forwarded to S. In this way, between S and D the routing information only partially is correct, but can not destroy among the source node and the destination node data packets the normal sending and receiving, the malicious nodes M₁ and M₂ collusion attack does not affect the SAWCMRP protocol performance.

(4) Black hole attack

The attacker in did not arrive at the destination node path, beat announced that there were at the node of the path, the source node set up through the malicious node routing, In subsequent messages are sent, malicious node abandon through its message, make source node and purpose nodes can't normal communication, so as to form a black hole attack. If the malicious node M₁ receives route request packet {QST S}; it produces false route reply packet {RST; S, M₁, D} to deceive the source node S. In the absence of security mechanism of AODV_D protocol, S will accept and use this route, because {S, M₁, D} look than the other route is short, and the M₁ and S very close, S will be the first to receive M₁. However, in SAWCMRP protocol, such response packet will be discarded, because M₁ not S and D shared key KSD, it can calculate the correct MAC.

8 Conclusion

This paper introduces the IKE protocol based on public key signature, its application in the wireless campus network Mesh the join new node and the access server and A node N jump range negotiate a shared key pair. Then describes the SAWCMRP protocol structure and design scheme, and in detail introduces the protocol routing algorithm. Finally, in view of the existing various wireless Mesh network attack method, for the security of SAWCMRP undertook comprehensive analysis.

References

1. Akyildiz, I., Wang, X., Wang, W.: Survey on wireless Mesh networks. *Computer Networks* 47(4), 445–487 (2005)
2. Hossain, E., Leung, K.K.: *Wireless Mesh Networks Architectures and Protocols*, pp. 34–67. Science Business Media, LLC, London (2008)
3. Perkins, C.E., Bhagwat, P.: Highly Dynamic Destination-sequenced Distance-Vector Routing (DSDV) for Mobile Computers. *Computer Communications Review*, 234–244 (1994)
4. Perkins, C.E., Royer, E.M.: Ad-hoc On-Distance Vector (AODV) Routing. In: *Proceedings of 2nd IEEE Workshop on Mobile Computing Systems and Applications*, pp. 78–86 (1999)
5. Chen, W.: IKE protocol analysis and implementation, pp. 16–20. Southwest Jiao Tong University, Chengdu (2005)
6. Li, X.: Message authentication code based On block cipher, pp. 31–34. Xi'an Electronic and Science University, Xi'an (2009)
7. Yan, Q.: Wireless Mesh network secure routing, pp. 40–42. Xi'an Electronic and Science University, Xi'an (2005)

Mobile Counteracted Time Synchronization for Underwater Sensor Networks

Ying Guo¹ and Yutao Liu²

¹ Qing Dao University of Science & Technology

² Qing Dao Yawei Software Co., Ltd. Qing Dao, China

qd.guoguo@gmail.com

Abstract. In energy constricted underwater wireless sensor networks (UWSN), time synchronization is a critical and challenging issue. Recently, numerous time synchronization protocols have been proposed for terrestrial wireless sensor networks, but they could not be directly applied to UWSN. It is because most of them do not consider about long propagation delay and sensor node mobility, which are important characteristics in UWSN. In this paper, we propose a Mobile Counteracted Time Synchronization approach, called “Mc-Sync”, which is a novel time synchronization scheme for mobile underwater acoustic sensor networks. It makes use of two mobile reference nodes to counteract the effect of node mobility. Analysis and simulation results show that Mc-Sync outperforms existing time synchronization schemes significantly.

Keywords: time synchronization, mobile reference node, node mobility, underwater sensor networks.

1 Introduction

Time synchronization is an important issue for many distributed applications, especially for sensor networks [1][2]. It requires collaboratively processing of time-sensitive data in the applications of environment supervision, target tracking and so on [3][4]. Recently many time synchronization protocols for terrestrial wireless sensor networks have been proposed [5][6]. As terrestrial communication is based on Radio Frequency (RF) technology, all of these mechanisms assume that propagation latency is negligible and can be effectively factored out of design consideration. While underwater communication mainly uses acoustic communication technology. There are several fundamental differences between RF communication and acoustic communication, such as large propagation delay and node mobility [7][8]. Thus protocols for terrestrial sensor networks could not be directly applied to underwater acoustic sensor networks.

Underwater sensor networks (UWSN) inherit many different features. Node mobility is one of the most important effects. From empirical observation, nodes without any self-propelling capability can move with wind and ocean current typically at the rate of 0.83-1.67m/s, and existing Autonomous Underwater Vehicles (AUV)

typically move at a rate up to $2.9m/s$. In addition, node movement also brings Doppler shift [9][10], which adds the difficulty to estimate propagation delay exactly. The propagation speed of underwater acoustic channel is five orders slower than radio waves, thus resulting in significantly longer propagation delay [11], which makes relatively large Doppler Effect and inter symbol interference. Moreover, UWSN has limited bandwidth, limited transmission rate and high bit error rate [12].

Recently, several underwater sensor network time synchronization algorithms have been proposed. TSHL [13] is the first time synchronization algorithm designed for high latency networks specifically. It uses one-way communication to estimate the clock skew and two-ways communication to estimate the clock offset. MU-Sync [9] runs two times of linear regression to estimate the clock skew and clock offset for cluster based UWSN. TSHL assumption of constant propagation delay, thus it can not handle mobility issues. MU-Sync assumes that the one-way propagation delay can be estimated as the average round trip time which causes extra errors, and has relative high message overhead.

Mobi-Sync [7] and D-sync [8] consider more about node mobility. Mobi-Sync need the help of surface buoys with GPS and synchronized super nodes. Ordinary nodes use a correlation model to estimate their velocity and launch time synchronization. D-sync exploits Doppler shift to provide an indication of the relative motion between nodes, and improves time synchronization precision. But they require specialty deployment or depend on the precision of velocity measurement, which is difficult to implement in underwater environment.

In this paper, we propose a mobile counteracted time synchronization algorithm for underwater acoustic sensor networks, called “Mc-Sync”. As UWSN experiences node mobility, our design utilizes two mobile reference nodes to eliminate this bad effect, and we examine the performance of Mc-Sync though simulations carefully.

The rest of this paper is organized as follows. In Section 2, we analyze the effect of node mobility and provide Mc-Sync algorithm in details. We present simulation results in Section 3 and offer our conclusion and future work in Section 4.

2 Algorithm Design

In this section, we analyze the effect of node mobility to the clock skew, and present design details of Mc-Sync, which makes use of two mobile reference nodes to counteract the impact of node mobility.

2.1 Effect of Node Mobility

In order to compute out the clock skew, making use of one-way information exchange twice is the simplest method. In this process, we assume that the propagation delays for each information exchange are the same. As shown in figure 1, reference node A with standard time is fixed, and node B is a sensor node to be synchronized. Node A sends synchronization information to node B. The message contains sending time T_1 , which is a time stamp of MAC layer. Node B receives this message and records the

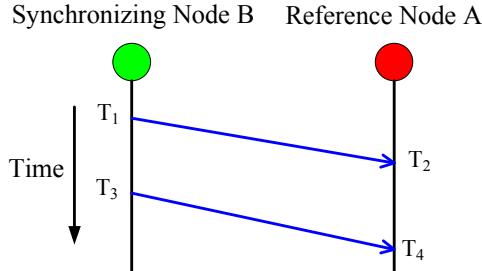


Fig. 1. Compute the Clock Skew

information receiving time T_2 with its local time. Then node A sends synchronization message again, which contains the sending time of MAC layer time stamp T_3 , node B receives this message and records the local receiving time T_4 .

Assuming that every information propagation process costs the same time t , we could get:

$$T_2 - t = aT_1 + b \quad (1)$$

$$T_4 - t = aT_3 + b \quad (2)$$

In which, a is the clock skew, b is the clock offset.

From equation (1) and (2), we could obtain:

$$a = \frac{T_2 - T_4}{T_1 - T_3} \quad (3)$$

Ideally, the clock skew is equal to the ratio of $T_2 - T_4$ and $T_1 - T_3$. However, in actual situation, a computed by this method is not only the clock skew, but the clock skew with time variety caused by node mobility. Due to long underwater propagation delay, effect of the clock skew is much less than the node mobility in the time synchronization process.

We give an example to illustrate above conclusion. Assume a equal to 1.00001, and the sensor node to be synchronized moves along the direction away from reference node. The node moves with maximum speed 2.9m/s. The time interval between two successive reference packets is 1 s. The average propagation speed in the simulated environment is 1500m/s with 1% fluctuates. Use equation (3), we could calculate that a is equal to 0.99982. That is to say, time synchronization makes the clock deviation bigger than before.

In order to solve this problem, and improve calculating precision of a , lots of existing algorithms adopt linear regression, which can not eliminate the influence of node mobility essentially. Some algorithms involved node position, speed and other information to computation process, but most of them require complex deployment and hard calculation. In order to reduce the effect of node mobility, we use two reference nodes to get exact clock skew.

2.2 Mc-Sync Design

As in figure 2, Node A1 and A2 are reference nodes with standard time, which located at opposite sides of node B and keep still. Node B is the node to be synchronized, which floats with ocean current.

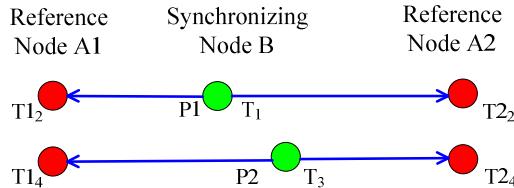


Fig. 2. Effect of Node Mobility

As in figure 2, Node B sends synchronization information at its local time T_1 , node A1 and A2 receive this information at time T_{12} and T_{22} respectively. We assume all the time used in time synchronization process is MAC layer time stamp. Node B sends synchronization information at its local time T_3 . Node A1 and A2 receive this information at T_{14} and T_{24} respectively. Due to the mobility of node B, it's information sending position changes from P1 to P2.

Adopt equation (3) to calculation a . There is a difference ΔT between actual ($T_{12}-T_{14}$)/($T_{22}-T_{24}$) and computed value, as the result of node mobility. The difference of A1 and A2 has the same value and opposite direction, as in figure 3.

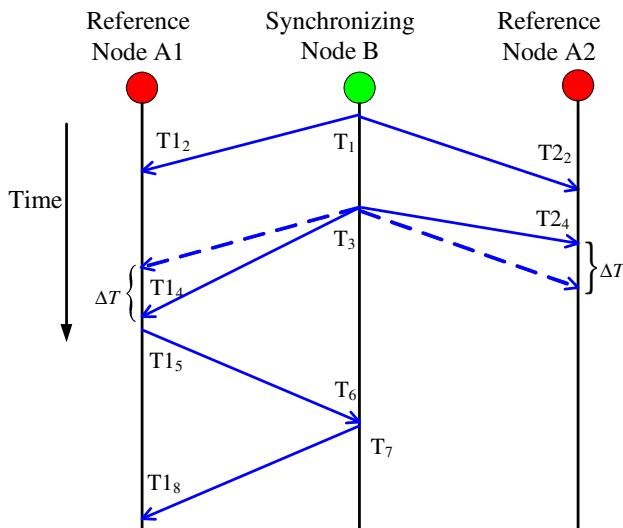


Fig. 3. Mobile Counteracted Time Synchronization

For node A1, we could obtain:

$$a = \frac{T1_2 - T1_4 - \Delta T}{T_1 - T_3} \quad (4)$$

In common, for node A2, we could obtain:

$$a = \frac{T2_2 - T2_4 + \Delta T}{T_1 - T_3} \quad (5)$$

From equation (4) and (5), we could compute out exact clock as follow:

$$a = \frac{T1_2 - T1_4 + T2_2 - T2_4}{2(T_1 - T_3)} \quad (6)$$

Then we use two-way information exchange to estimate the clock offset, as in figure 3. Node A1 transmits synchronization message to node B, after it has received the second synchronization packet. The message includes MAC layer time stamp $T1_5$. Node B receives this packet, records the receiving local time T_6 and immediately sends synchronization information at time T_7 . Node A1 receives this message and records the receiving time $T1_8$. As node B immediately replies to node A1, the time interval is very short, node B is barely moving. So there is scarcely any error of the clock offset caused by node mobility. We could get:

$$T1_5 + t = aT_6 + b \quad (7)$$

$$T1_8 - t = aT_7 + b \quad (8)$$

From equation (7) and (8), we could compute out the clock offset:

$$b = \frac{T1_8 + T1_5 - a(T_7 + T_6)}{2} \quad (9)$$

As discussed above, we could get sensor node's clock skew and offset from equation (6) and (9).

3 Simulations

We analyze the preferences of Mc-Sync via simulations, and compare it with No-Sync, TSHL [13] and MU-Sync [9] in this section.

In our simulations, we assume two reference nodes have standard time. And the sensor node to be synchronized has its own inside clock, which clock skew is $8ppm$ and clock offset is $10ppm$. It moves in the deployment area with maximum speed (V_{max}) $2.9m/s$, and changes its speed randomly within the range of $[0, V_{max}]$. The time interval between two successive reference packets is $1 s$. Clock granularity is $1\mu s$. Receive jitter is $15\mu s$. The time stamps in simulations are the time stamp of MAC

layer. The average propagation speed in simulated environment is $1500m/s$ with 1% fluctuates. The number of reference packets used by TSHL and MU-Sync to perform linear regression is 20.

In the first simulation, we research on the error with time elapsed since synchronization and compare Mc-Sync with different algorithm, e.g. MU-Sync, TSHL and No-Sync. As in figure 4, with time goes by, errors grow after time synchronization. As the clock skew leads to error increasing after synchronization, this comparison result actually demonstrates different accuracies on the clock skew these algorithms can achieve. In this simulation, the error for Mc-Sync is 27.4% of TSHL, and 60.9% of MU-Sync. The high precision of Mc-Sync arises from it takes node mobility into consideration. It utilizes two reference nodes to counteract the error caused by node mobility in time synchronization process.

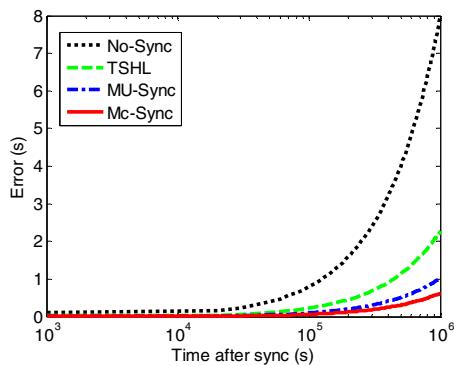


Fig. 4. Error & Time since Synchronization

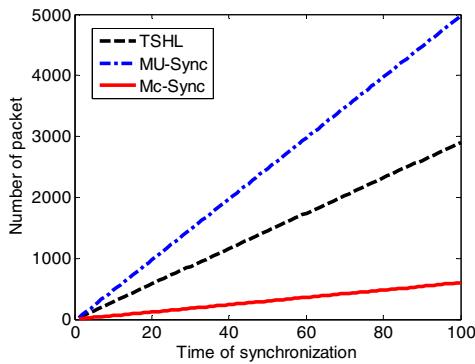


Fig. 5. Number of Packet

Then we study the energy cost of these algorithms. Figure 5 shows the number of packets needed in synchronization process, which represents the energy cost of TSHL, MU-Sync and Mc-Sync. As in figure 5, the packet number of Mc-Sync is the smallest compare with TSHL and MU-Sync.

MU-Sync and TSHL have higher packet number as they run linear regression to estimate the clock skew and clock offset. Mc-Sync only adopts five times of message exchange, and does not use linear regression, which reduces the energy cost greatly. As we can see, the packet numbers of Mc-Sync is 20.7% of TSHL, and 12.1% of MU-Sync.

As node mobility is one of the main effects to time synchronization, we also discuss the effect of node velocity in our simulation. We change the node speed from 0m/s to 10m/s . Figure 6 shows the impact of sensor node velocity compared Mc-Sync with TSHL and MU-Sync. The error of Mc-Sync is 11.5% of TSHL, and 17.9% of MU-Sync.

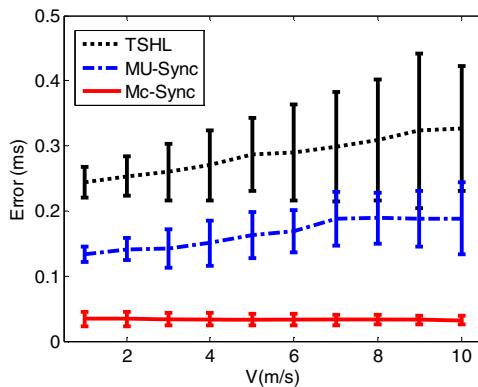


Fig. 6. Effect of Velocity

As in figure 6, with the parameter velocity's increase, the synchronization errors of TSHL and MU-Sync increase faster than Mc-Sync. This is because both TSHL and MU-Sync do not consider about propagation delay caused by node mobility. As Mc-Sync counteracts the effect of node mobility, its synchronization error is barely affected by velocity increase. As a result, Mc-Sync performs better than No-Sync, TSHL and MU-Sync significantly.

4 Conclusion

In underwater sensor networks, node mobility has significant influence on the performance of time synchronization protocols. We research on the effect of node mobility, and make use of two mobile reference nodes to counteract the impact of node mobility. We present Mc-Sync, and analyze its preference though simulation. In our future work, we plan to examine the applicability of our algorithm in more complex underwater environments.

Acknowledgments. This work is supported by the NSF China grant No. 61103196 and 61170258. Research Award Fund for Excellent Young Scientist of Shandong Province No. BS2012DX011. Research Funds Provided to New Recruitments of Qing Dao University of Science & Technology No. 0022465.

References

1. Heidemann, J., Ye, W., Wills, J., et al.: Research Challenges and Applications for Underwater Sensor Networking. In: IEEE Wireless Communication and Networking Conference, Washington, pp. 228–235 (2006)
2. Wang, Y., Zhang, Z., Bi, G.: An Overview on Underwater Acoustic Sensor Networks. In: 17th International Conference on Telecommunications, pp. 779–783 (2010)
3. Liu, L., Xiao, Y., Zhang, J.: A Linear Time Synchronization Algorithm for Underwater Wireless Sensor Networks. In: IEEE International Conference on Communications (ICC 2009), pp. 14–18 (June 2009)
4. Zhou, Z., Cui, J.H., Bagtzoglou, A.: Scalable Localization with Mobility Prediction for Underwater Sensor Networks. In: Proceedings of IEEE INFOCOM 2008, Mini-Conference, Phoenix, Arizona, USA (2008)
5. Lasassmeh, S.M., Conrad, J.M.: Time Synchronization in Wireless Sensor Networks: A survey. IEEE Southeast Con., pp. 242–245 (2010)
6. Wu, Y.-C., Chaudhari, Q., Serpedin, E.: Clock Synchronization of Wireless Sensor Networks. IEEE Signal Processing Magazine 1(8), 124–138 (2011)
7. Liu, J., Zhou, Z., Peng, Z., Cui, J.-H.: Mobi-Sync: Efficient Time Synchronization for Mobile Underwater Sensor Networks. In: IEEE GLOBECOM, pp. 1–5 (March 2010)
8. Lu, F., Mirza, D., Schurger, C.: D-Sync: Doppler-Based Time Synchronization for Mobile Underwater Sensor Networks. In: The Fifth ACM International Workshop on Underwater Networks (WUWNET), Woods Hole, MA (2010)
9. Chirdchoo, N., Soh, W.S., Chua, K.C.: MU-Sync: A Time Synchronization Protocol for Underwater Mobile Networks. In: Proceedings of WUWNet, pp. 35–42 (2008)
10. Parrish, N., Roy, S., Arabshahi, P.: Symbol by Symbol Doppler Rate Estimation for Highly Mobile Underwater OFDM. In: Proceedings of the Fourth ACM International Workshop on Underwater Networks (WUWNet 2009), New York, NY, USA, pp. 1–8 (2009)
11. Zhou, Z., Peng, Z., Cui, J.-H., Shi, Z.: Efficient Multipath Communication for Time-Critical Applications in Underwater Acoustic Sensor Networks. IEEE/ACM Transactions on Networking, 28–41 (August 5, 2010)
12. Ismail Nor-Syahidatul, N., Abdullahi, H.L., Ariffin Sharifah, H.S.: Analyzing the Performance of Acoustic Channel in Underwater Wireless Sensor Network (UWSN). In: Asia Modelling Symposium, pp. 550–555 (2010)
13. Syed, A.A., Heidemann, J.: Time Synchronization for High Latency Acoustic Networks. In: 25th IEEE International Conference on Computer Communications (INFOCOM), pp. 1–12 (2006)

Cross-Layer Routing Metric for Wireless Mesh Networks

Xiang Li, Yong-Jian Yang, Chun-Sheng Cui, and Guan-Nan Yang

Computer Science and Technology of Jilin University,
Communication Software and Agreement Engineering Laboratory
15843043878@163.com

Abstract. To meet the shortage of the typically routing metrics used in wireless mesh network, this paper proposed a new routing metric for wireless mesh network—CAETT (Congestion Avoidance Expected Transmission Time), comparing with several other routing metrics. With the queue's utilization rate and the transmit situation of control frames in 802.11 protocol's MAC layer, a reasonable path is chosen in terms of the channel competition status, link data frame delivery rate and the node's queue utilization rate. After that the advantages of this new routing metric and its applications were discussed. Finally the simulation experiments show the good network performance.

Keywords: wireless mesh network, routing metric, channel competition, expected transmission time, queue utilization rate.

1 Introduction

Wireless Mesh Network also called wireless multi-hop network; it is a new type of broadband wireless network structure. WMN can be regarded as the wireless local area network and the mobile AD-HOC network integration; it's the Internet's wireless version [1]. Routing problem is always a research focus in wireless mesh network. Because of the bad wireless condition, channel instability and interference in wireless network, it's still a challenge to make effective routing design [2]. A good routing metric should be able to accurately reflect the characteristics of the link, ensure short access time, short transmit delay, high data delivery rate etc.

In the reference 3 the author presents several common routing metrics in wireless mesh network, and analyzes the routing algorithms. The current status of the design scheme of cross-layer routing considering the features of the wireless link, a routing metric usually consists of a set of measures that are mathematically combined to pick up the link quality. In the reference 4 Vinicius etc proposed that the routing metrics for WMN have followed four main trends, which are: basic, interference-aware, load-aware and a combination of interference-aware and load-aware [4].

The most recent and relevant routing metrics for WMN includes ETX, ETT, WCETT, MIC, WCETT-LB etc [4]. ETX (expected transmission count) defines the expected number of MAC layer transmissions for deliver a packet through a wireless link successfully, including retransmission. The data package delivery ratios and the ACK package delivery ratios are measured with the aid of broadcast probe packets.

ETX doesn't consider the transmission rate, and cannot provide load balancing. It just indirectly describes the network interference, doesn't consider the overhead of channel competition. Also the active probing method used in this metric introduces inaccuracies on the estimation of the loss rate measure, because of its difference with data packet. ETT (expected transmission time) extends ETX by taking account of the differences in link transmission rates. WCETT[5] (weighted cumulative ETT) routing metric extends ETT to consider the intra-flow interference; it adds the link ETT that use the same channel in the path together, and picks the maximum sum as the intra-interference metric. WCETT has the same drawbacks of ETT. It does not take account of the effect of traffic loads and this may lead to paths through congested areas. What's more, WCETT lacks an isotonicity property, and is pessimistic for longer paths [4]. MIC (metric of interference and channel switching) [6] also extends ETT to estimate the inter-flow and intra-flow interference. It uses channel switching cost to depict the intra-flow interference and interference-aware resource usage to depict the inter-flow interference. To ensure its isotonicity property, MIC employs a strategy that introduces a virtual network to calculate channel switching cost which leads it into a complex algorithm. MIC does not take account of traffic load measures. WCETT-LB (Weighted Cumulative ETT-Load Balancing) [5] takes account of the traffic load and logical interference that are not captured by WCETT. However, WCETT is not isotonic and does not detect interference in a dynamic way.

In view of the features and insufficient of the above routing metrics, this paper presents a new routing metric that comprehensive considers of channel competition state, link date frame delivery rate and the queue congestion degree of the node—Congestion Avoidance ETT (CAETT).

2 Cross-Layer Routing Metric

2.1 The Advantage of Calculate Routing Metric in Downstream Node

Suppose that a node has n neighbor nodes, when the node initiates a routing request, the forms of RREQ messages are show in table 1 and table 2.

Table 1. The RREQ for calculate routing metric in downstream node

RREQ ID	sourceAddr	destAddr	Route record	Total routing metric
---------	------------	----------	--------------	----------------------

Table 2. The RREQ for calculate routing metric in upstream node

REEQ ID	Source-Addr	DestAddr	Route record	The routing metric to neighbor 1	...	The routing metric to neighbor n
---------	-------------	----------	--------------	----------------------------------	-----	----------------------------------

Generally, the routing metric is calculated in the upstream node [7]. During the routing request of on-demand protocols, the upstream node needs to add all the paths' routing metrics into the RREQ message. This will greatly increase the length of

RREQ message, and make its length very unstable (the length is determined by the number of the node's neighbor), as shown in table 2. So we calculate the routing metrics in the downstream node. The RREQ message only need to bring the route record and total routing metric, when the downstream node receive the RREQ message, it calculate the new CAETT according to the link metrics it keeps. The link metric shows the link quality between its upstream node and itself. We let the downstream node calculate the metric of links between its upstream nodes and itself and consider the downstream node's queue usage. Because all the information is stored in the downstream node, we calculate in the downstream node and keep the RREQ message short and fixed without increasing the cost of the information interaction.

2.2 Link Expected Transmission Time

In the 802.11, MAC layer use the RTS/CTS mechanism to compete channel. In order to describe the link condition between two neighbor nodes, we use the passive measurement to get link expected transmission time. On the one hand it doesn't need to import active probing packets, can reduce the costs; on the other hand the measurements is according to the general communication, there won't be imprecision problems caused by the differ between probing packets and data packets [4].

The two nodes communication process can be represented as follows: node A broadcasts RTS asking for access the channel between A and B, if B receives A's RTS and the channel is idle, it broadcasts CTS to reply A's request and inform other nodes that the channel is unavailable. Then A send data packet to B, and B reply an ACK when receives the data correctly. So the expected transmission time in the link can be defined as:

$$ETT = T_{col} + T \tan \quad (1)$$

Where T_{col} is the collision time that needed for channel competition, and $T \tan$ is the data transmission time.

According to the content in table 1, we do the calculation in the downstream node. The downstream node saves the RTS sending times in the latest communication during channel competition. Add a new field N in the RTS to record the times. When the upstream node receives a CTS frame successfully, set N 0; else once a RTS was sent add 1 to N. As shown in the fig. 1: in the last RTS frame, N=4.

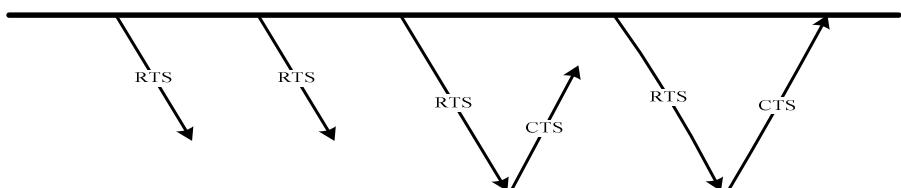


Fig. 1.

The downstream node calculates the channel competition time by the N in the recently received RTS frame.

$$T_{col} = N \times (Trts + \tau) \quad (2)$$

Where $Trts$ represents the sending delays of a RTS frame and τ is the back off time.

After the node compete a channel, it needs to transmit the data message. The downstream node uses two counters: $Nrts$ and $Ndata$ to record the RTS and data frames that are successfully received in an average period from an upstream node. The $Nrts$ represents the number of date frames the upstream node prepared to send to the downstream node in the average period, while the $Ndata$ represents the number of data frames that the downstream node actually received. Hence the data expected transmission count in the link can be defined as formula 3:

$$ETX_{data} = \left(\frac{Nrts}{Ndata} \right) \times 100\% \quad (3)$$

The ETX can reflect the link transmit quality and interference in the network very well. It is a description of link date frame delivery rate. According to the ETX, we can express the data transmission time as follows:

$$T_{tan} = ETX_{data} \times (T_{data} + T_{cts}) + T_{ack} \quad (4)$$

Where T_{data} , T_{cts} and T_{ack} respectively represent the sending delays of data frame, CTS frame and ACK frame.

To sum up, according to the formula 2 and 4 we can define the link expected transmission time as $ETT = T_{col} + T_{tan}$

2.3 Queue Utilization Rate

The queue utilization rate of a node can be express as QUR (queue utilization rate).

$$QUR = \frac{U}{B} \quad (5)$$

Where U is the part that has been used in the queue and B is the total queue capacity of a node. The smaller the QUR is, the idler the node is. From the view of fair, this part of “free” nodes should be involved in the routing. In addition, when the queue utilization rate of a node is high, it may cause congestion, packets loss etc which will increase transmission delay and affect the network throughput.

2.4 Routing Metric—CAETT (Congestion Avoidance Expected Transmission Time)

In comprehensive consideration of the link expected transmission time and queue utilization rate of the nodes. According to the content above, the CAETT of link i can be defined as:

$$CAETTi = ETTi + QURi \times \frac{Sb}{Bd} \quad (6)$$

Where Sb is the total size of the queue and Bd is the transmission rate at a node. $ETTi$ is the expected transmission time of link i, including channel competition time and data transmission time. $QURi \times \frac{Sb}{Bd}$ represents the queuing time in the downstream node, it is used to depict the processing ability and congestion degree of the node.

Then, for a route p, CAETT is defined as:

$$CAETT = \sum_{i \in link_p}^n CAETTi \quad (7)$$

CAETT uses the channel competition time and link data expected transmission time to describe channel busy status and link quality directly and the interference indirectly. And take account of load-balancing by considering the downstream node's queue utilization rate. It depicts link delays, interference and traffic load together. Hence use this routing metric can keep away from congestion links and nodes, choose the path with shortest transmission delay. CAETT can effectively reduce the end-to-end delay, improve group delivery rate and network throughput.

3 Cross-Layer Routing Design

We explain the congestion avoidance expected transmission time by DSR protocol.

Considering that using 802.11 in the MAC layer, we add a new filed in the RTS frame named N, which is used to record the send times for the RTS until the node gets a channel. Each node maintains a structure: including sourceAddr, N and T. N means channel competition time, T is used to represent data expected transmission time

The downstream node gets the sourceAddr and N from the RTS frame, and updates the N through the corresponding sourceAddr. Also the node needs to periodically calculate T to ensure the downstream node can get the link metrics at any time.

$$T = ETXdata \times (Tdata + Tcts) + Tack$$

The combination of the routing metric and DSR protocol:

The main principle to choose path: Use CAETT to replace hop count as routing metric.

How to get CAETT: When a node gets a RREQ; it will recognize the upstream node by the route record. Then it takes out the N and T from structure LinkQuality by the corresponding sourceAddr, and calculates the CAETTi by the instant queue utilization rate (BUR) of the node. CAETTi is the metric of the link between the upstream node and itself.

$$CAETTi = N \times (Trts + \tau) + T + QUR \times \frac{Sb}{Bd}$$

The use of CAETT:

1. If the destination node receives more than one RREQ, it updates the total routing metric (CAETT) first, then choose the path with the smallest CAETT.
2. If the intermediate node receives some RREQs that have the same source address and destination address. Then it has to update the CAETT first, and compare it with former CAETT. If the former is smaller, than drop the RREQ; else record the CAETT and continue broadcast this new RREQ. If the intermediate node has a path to the destination, it replies the RREQ immediately.

4 Network Simulation and Analysis

We use Glomosim [10] to simulate the network performance. In the MAC layer, we use 802.11, and slightly revise the control frame. The frequency in physical layer is 2.4GHz, while the bandwidth is 2Mb/s, the communication distance is 265m [9].

Let 30 nodes distribute in the range of 800mx800m in a random way. In view of the standard static characteristic of wireless mesh network, we set the mobility of the nodes none. And set up 9 random CBR application sessions. Each CBR session sends 30 messages in the same rate. The message size is 512 bytes. By adjusting the sending interval between two messages, we make statistics of the packet delivery rate, network throughput and average end-to-end delay in different network loads [8]. The results are show in fig.2, 3 and fig. 4.

According to the simulation results, when the network loads is low, the differences between the two routing metrics are small. Along with the gradual increase in sending rate, the network loads increase, the performance of CAETT becomes better than hop count in every aspect. The main reason is that using hop count as routing metric cannot take account of the link state and the node's processing ability. When the network loads reaching saturation, it may cause node's congestion and greater channel competition which aggravate end-to-end delay and decrease the network throughput. The CAETT as we proposed, comprehensive considers of channel competition, link quality, and load-balancing. So it can avoid "hot" nodes, and choose

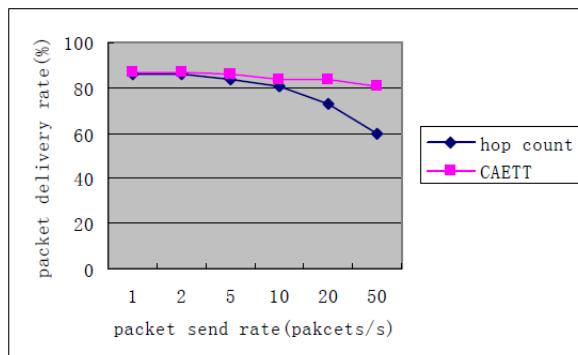


Fig. 2. Packet delivery rate

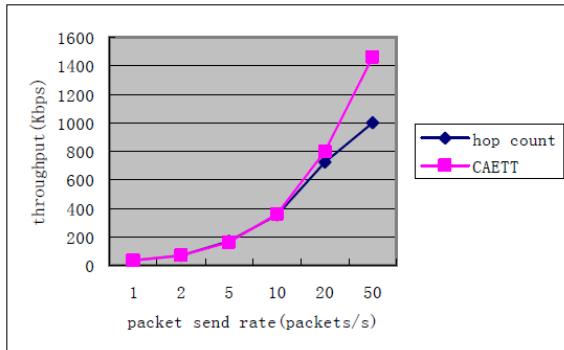


Fig. 3. Total network throughput

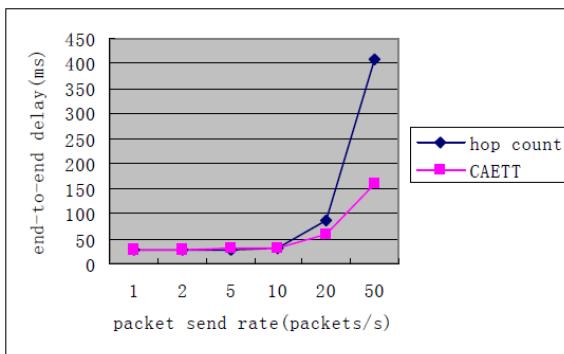


Fig. 4. Average end-to-end delay

a path with good link quality and low load. Avoid package loss due to excessive channel competition and node's congestion. So that this routing metric can reduce the end-to-end delay, and increase the packet delivery rate and network throughput effectively.

5 Summary

According to the standard static characteristic of wireless mesh network, this paper proposed a new cross-layer routing metric Congestion Avoidance Expected Transmission Time which comprehensive considers of channel competition state, link data delivery rate and the queue utilization rate of the node. This new metric can reflect instantaneous channel situation, node's load directly, and according to the channel status and queue utilization rate to select the link of good quality and involved relatively free node in transmission. At the same time it can consider the impact of network interference indirectly. Simulation result shows that using the routing metric proposed in this paper can effectively increase the packet delivery rate and the network throughput when the network is in high loads. Also reduce the end-to-end delay to a certain extent.

References

1. Akyildiz, I.F., Wang, X., Wang, W.: Wireless mesh networks: a survey. *Computer Networks* 47, 445–487 (2005)
2. Hossain, E., Leung, K.K., et al.: *Wireless Mesh Networks Architectures and Protocols*, vol. 5, pp. 1–22, 69–97. China Machine Press, Beijing (2009)
3. Wang, X.-K., He, D.-P.: A Research on Wireless Mesh Network Routing Metric. *Date Communications* (4), 30–33 (2008)
4. Borges, V.C.M., Curado, M., Monteiro, E.: Cross-layer routing metrics for mesh networks: Current status and research directions. *Computer Communications* 34, 681–703 (2011)
5. Ma, L., Denko, M.: A routing metric for load-balancing in wireless mesh networks. In: *21st International Conference on Advanced Information Networking and Applications Workshops, AINAW 2007*, vol. 2, pp. 409–414 (2007)
6. Yang, Y., Wang, J., Kravets, R.: Interference-aware Load Balancing for Multihop Wireless Networks. Tech. Rep. UIUCDCS-R-2005-2526. Department of Computer Science, University of Illinois at Urbana-Champaign (2005)
7. Shao, Y., Su, J., Xiao, M.: Routing Protocol of Wireless Mesh Networks Based on Cross-layer. *Modern Electronics Technique* (19), 27–29 (2009)
8. Shen, C., Lu, Y.-F., Xia, Q.: An integrated metrics based routing protocol for wireless mesh networks. *Chinese Journal of Computers* 33(12), 2300–2311 (2010)
9. Zhu, Y.-F., Lei, L., Xu, Z.-Z.: Cross-layer Routing Protocol for Ad Hoc Networks Based on Synthetical Metric. *Computer Engineering* (18), 79–81 (2010)
10. Gao, Z.-G.: *Glomosim Network Simulation—CALMSANNY*. Post and Telecom Press, Beijing (2003)

Optimization and Realization of Ad Hoc On-Demand Distance Vector Routing Based on NS2

Deyu Lin and Xiaojing Meng

College of Information Science and Technology, Shandong University of Science and Technology, No.579, Qianwangang Road, Economic & Technical Development Zone, Qingdao, Shandong, China
1dy19880210@126.com, jack.meng@263.net

Abstract. In the standard AODV protocol, the routing table keeps a relatively fixed field called lifetime to judge whether a route table entry is overdue or not. In this way, the routing table can be renewed dynamically. However, this method doesn't take the request number of a node into consideration, which will do harm to the improvement of the network performance. In view of this case, the author came up with an improved protocol called MYAODV which used the request number to determine the lifetime of each table route entry, and then made it come true based on NS2. To verify the effect of this new protocol, a series of experiments were made. The result of the experiments shows that the new protocol can truly make the network perform better.

Keywords: routing protocol, NS2, AODV, MYAODV, network performance.

1 Introduction

Mobile Ad Hoc network (MANET) is an autonomous system of mobile nodes connected by wireless links. In this system, each node has an equal status and without any center entities. Each node operates not only as an end system, but also as a router to forward packets. As a result, the adaptability of this network is very strong. So far, there have been two kinds of Ad Hoc routing protocols. One is proactive routing protocol which is also called “table-driven” routing protocol. By broadcasting routing message periodically to the nodes nearby, each node updates the routing table to keep the routing message consistently. Wireless Routing Protocol (WRP) and the Destination Sequence Distance Vector (DSDV) are the typical proactive routing protocols. The other one is reactive routing protocol which is also called “demand-driven” routing protocol. Compared with the proactive routing protocol, less control overhead is a distinct advantage of the reactive routing protocol. The Dynamic Source Routing (DSR) and Ad Hoc On-demand Distance Vector routing (AODV) are examples for reactive routing protocols. Recently, the AODV protocol is the widely used wireless network protocol.

2 AODV Protocol

2.1 AODV Protocol Profile

Ad Hoc On-Demand Distance Vector (AODV) routing is a routing protocol for mobile Ad Hoc networks and other wireless Ad Hoc networks. It is an on-demand and distance-vector routing protocol, meaning that a route is established by AODV from a destination on demand.

The AODV protocol consists of two important mechanisms, route discovery and route maintenance. It uses four kinds of messages to achieve the functions of route discovery and route maintenance. They are Route Request Message (RREQ), Route Reply Message (RREP), Route Error Message (RERR) and Hello Message.

2.2 The Process of Route Discovery

A node disseminates a RREQ when it determines that it needs a route to a destination and does not have one available. The node that receives the RREQ first creates or updates a route to the previous hop without a valid sequence number. It then checks to determine whether it has received a RREQ with the same ¹Originator IP Address and ²RREQ ID within at least the last ³PATH_DISCOVERY_TIME. If such a RREQ has been received, the node silently discards the newly received RREQ, otherwise it replies with a RREP message. Once the RREP received by the source node the sends the RREQ message, the process of route discovery completes and the routing information is recorded.

2.3 The Process of Route Maintenance

A node may offer connectivity information by broadcasting local Hello messages periodically. If, within the past ⁴DELETE_PERIOD, it has received a Hello message from a neighbor, and then for that neighbor does not receive any packets (Hello messages or otherwise) for more than ⁵ALLOWED_HELLO_LOSS * HELLO_INTERVAL milliseconds, the node should assume that the link to this neighbor is currently lost. When links break, AODV causes the affected set of nodes to be notified so that they are able to invalidate the routes using the lost link and then a routing error is passed back to a transmitting node and finally some measures should be taken to repair the link, and the process repeats.

¹ The IP address of the node which originated the Route Request.

² A sequence number uniquely identifying the particular RREQ when taken in conjunction with the originating node's IP address.

³ The time of a route discovery process.

⁴ This is the time that tells a node when to delete a invalid route entry.

⁵ This time is added when whenever a node receives a Hello message from a neighbor node.

3 The Shortcomings of the AODV Protocol

In the standard AODV protocol, whenever a destination node or an intermediate node disseminates a RREP, it copies the value ⁶MY_ROUTE_TIMEOUT or the difference between the expiration and the current time into the field lifetime of the RREP. It is clear that this approach doesn't take the number of requests of a node into consideration, which causes some unnecessary overhead and thus is bad for the comprehensive performance of Ad Hoc network.

Let us consider such a situation: in the following fig.1, it is a simple network with three nodes, if node 2 frequently communicates with node 0, while node 1 seldom does. In the standard AODV protocol, the node 2 and node 1 have the same lifetime by the feedback from node 0. In this case, the performance of the network will be more or less affected by this fixed lifetime.

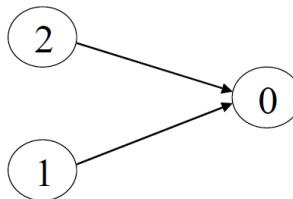


Fig. 1. A simple network with three nodes

4 Optimized Scheme

In view of the above problems, the solution can be as follow: adding a field of requests number to each route table entry, in the meanwhile, keeping a counter for each node. So whenever a node receives a RREQ, it increases its counter if it is the destination, or it increases the field which keeps the number of requests in its route table entry. When generating a RREP message, a node will make the lifetime related to the field of requests number (if it is the intermediate node) or the counter (if it the destination node). In this way, the lifetime of each route table entry can increase along with the requests times of a node. Thus the network delay, delay jitter and packet loss rate will somewhat decrease.

5 Realization and Verification

5.1 Realization

In order to achieve the above scheme, a new field named rt_dstcnt should be added to each route table entry to represent the number of requests of every destination node.

⁶ This is the time copied into the lifetime field when a destination node receives a RREQ message.

Given sometimes the node that receives the RREQ message is just the destination node, another field called cnt should also be added. Thus if a node receives a request, it will increase its cnt and reply with a RREP message. In the meanwhile, a new field named rp_dstcnt should be added to each RREP message to give a feedback to the node that sends the RREQ message so that the lifetime will increase along with link number dynamically.

Once receiving a RREQ message, a node will do the following operation:

```
cnt=(cnt+1)%20; //If it is the destination node
rt->rt_dstcnt=(rt->rt_dstcnt+1)%20;//If it is the
//Intermediate node
```

Then it replies with a RREP message, one more parameter rt->rt_dstcnt should be added to the function senReply to determine the lifetime of each route:

```
rp->rp_lifetime = lifetime*(1+rp->rp_dstcnt/60.0);
```

When a node uses the HELLO message to achieve the process of route maintenance:

```
nb->nb_expire = CURRENT_TIME
+(1.5*ALLOWED_HELLO_LOSSHELLO_INTERVAL)*(1+rt-
>rt_dstcnt/60.0);
```

Finally, in the ns-allinone-2.30/ns-2.30 directory, I created a folder named myaodv, then wrote the following source programs: myaodv_rtable.h, myaodv_rtable.cc, myaodv_packte.h, myaodv_rqueue.cc, myaodv.h, myaodv.cc, myaodv_logs.cc and myaodv.tcl. After finishing all these, I recompiled the NS2 and wrote a tcl script file for verification.

5.2 Experimental Verification

In this paper, I used a simulation software named NS2 to simulate the AODV and MYAODV protocols, and compared the mean network delay, mean delay jitter and packet loss rate when using the different protocols. Furthermore I used Gnuplot to graphically represent these three performance parameters for both protocols and hence analyzing them.

The network model includes eleven nodes which distributed in a square field of 1000m×1000m. Every node used the same configuration except the routing protocol, the time of simulation lasted 90 seconds.

When analyzing the network performance through the script, usually we compare the performance curve in the different data transmitting rate. In this paper, the data transmission rate ranged from 50Kb/s to 220Kb/s. I compared the difference of the performance between the AODV and MYAODV protocols.

The fig.2, fig.3 and fig.4 show the comparison of the packet delay, delay jitter and packet loss rate when using different protocols. It is clear that the network delay, delay jitter and packet loss rate decrease if using the MYAODV protocol. It is

because of the record of the number of requests in every route table entry that the lifetime of each route table entry can be extended.

To summarize what has been mentioned above, it is clear the performance of Ad Hoc will be improved based on the MYAOV protocol.

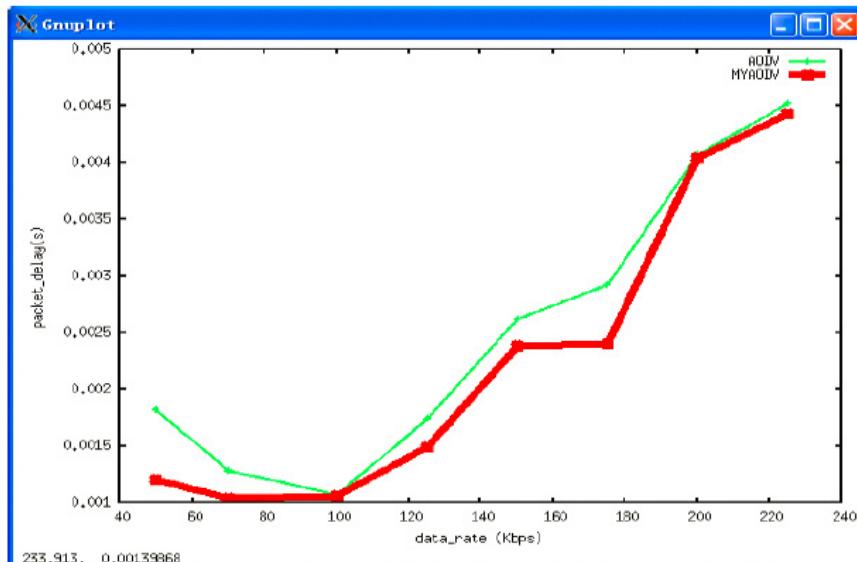


Fig. 2. Network delay

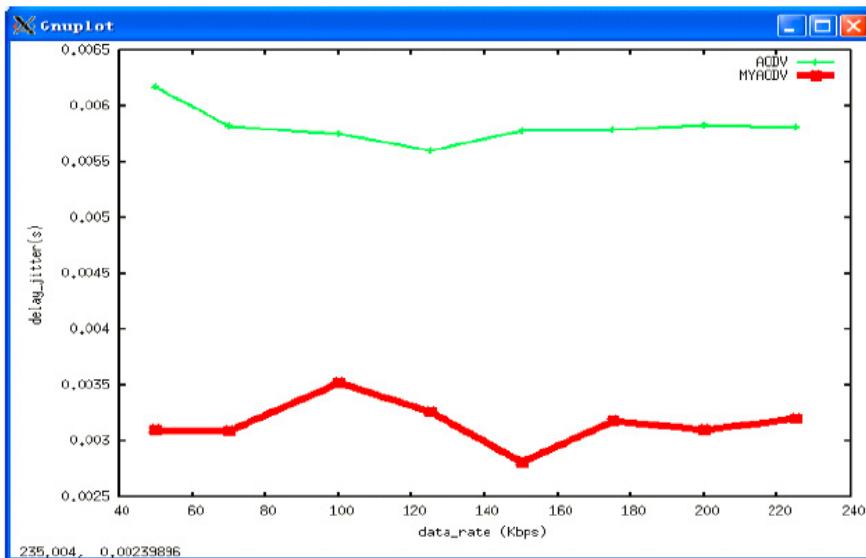
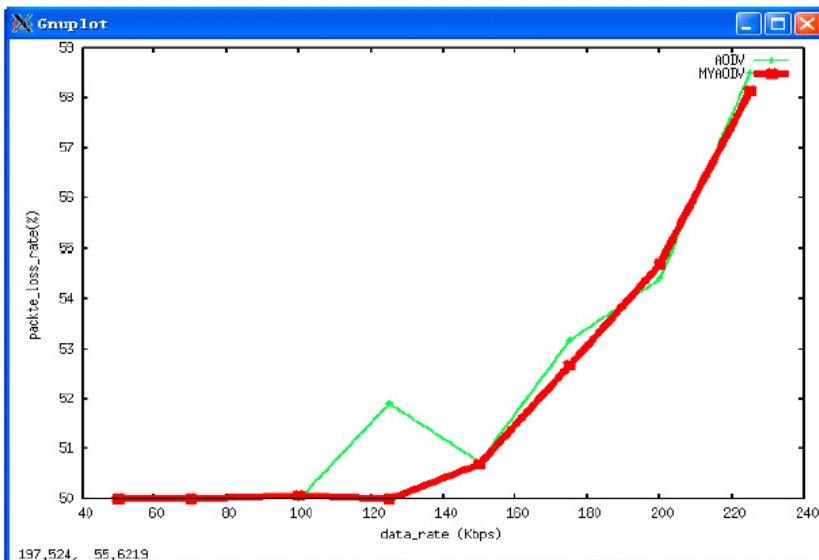


Fig. 3. Delay jitter

**Fig. 4.** Packet loss rate

6 Conclusion

This paper firstly pointed out the shortcoming of the standard AODV protocol and then came up with an optimized scheme, finally made it come true and designed an experiment to verify this scheme based on NS2. The result of the experiment proved that this scheme makes for the improvement of the Ad Hoc network. It will reduce the Network delay, Network delay jitter and packet loss rate. This can make the picture more smoothly and reduce the mosaic if used in IPTV.

Due to personal knowledge and information Limited, there may be some shortcomings in this paper. In order to make the AODV protocol have a better value of practical application in the future, much more energy should be spent on the security, Qos and multicast protocol which are waiting for further research.

References

1. Goel, A., Sharma, A.: Performance Analysis of Mobile Ad-hoc Network Using AODV Protocol. International Journal of Computer Science and Security (IJCSS) 3, 334–343 (2009)
2. Husieen, N.A., Ghazali, O.B., Hassan, S., Kadhum, M.M.: The Effect of Pause Time on the Performance of Mobile Ad-hoc Network Routing Protocols. In: 4th International Conference on Intelligent Information Technology Application, pp. 412–415. IEEE Press, New York (2010)
3. The Network Simulator ns-2,
<http://www.isi.edu/nsnam/ns/ns-documentation.html>

4. Perkins, C., Belding-Royer, E.: Ad Hoc On-Demand Distance Vector (AODV) Routing. IETF RFC 3561 (2003)
5. Wang, H., Luo, Z.: Research and Improvement of NS2-based AODV Protocol in Ad Hoc Networks. Computer Era (5), 12–14 (2011)
6. Saad, M.I., Ahmad, Z.: Performance Analysis of Random-Based Mobility Models in MANET Routing Protocol. European Journal of Science Research, 444–454 (2009) ISSN 1450-216X. 32
7. Zhou, Z., Li, Y., Zhou, Z., Chen, H.: Research of route protocol of based on Ad Hoc network. Journal of Central South University of Forestry & Technology 30, 156–162 (2010)
8. Zhang, R., Hui, X., Jiang, L.: AODV performance optimization simulation. Journal of Ningbo Polytechnic 14, 42–45 (2010)
9. Sargolzaey, H., Moghanjoughi, A.A., Khatum, S.: A Review and Comparison of Reliable Unicast Routing Protocols For Mobile Ad Hoc Networks. International Journal of Computer Science and Network Security (IJCSNS) 9, 186–196 (2009)
10. Yang, B., Li, Q.: Analysis and Improvement of AODV routing Protocol in MANET. Computer Science 37(7A), 93–95 (2010)
11. Yu, Z.: An Kind of AODV Protocol based on Energy improved. Modern Computer 1, 25–29 (2009)
12. Singh, J., Singh, P., Rani, S.: Enhanced Local Repair AODV (ELRAODV). In: International Conference on Advances in Computing, Control, and Telecommunication Technologies (2009)

A Double-Links Scheme for DSDV in Mobile Ad Hoc Networks*

Jia Lu, Junhui Wang, Gang Han, Baoliang Li, and Wenhua Dou

School of Computer, National University of Defense Technology Changsha,
Hunan, P.R. China
lujia661@126.com

Abstract. To deal with invalid route reconstruction problems in Destination Sequenced Distance Vector (DSDV) routing protocol, a Double-Links scheme for DSDV, DLDSDV is proposed. The most distinguished characteristic of DLDSDV is that it needs not any new message exchange for invalid route reconstruction. However, it has to use more memory to store an extra routing table, the secondary routing table. When a node detects a link is broken, the node in DLDSDV employs the secondary routing table to find the neighbor quickly which has a valid route to the destination. We extend the path availability model for MANETs and derive an expression of path holding probability of DLDSDV. Simulation results show that DLDSDV improves the packet delivery ratio and has almost the same routing overhead and end-to-end delay as DSDV.

Keywords: DSDV, route reconstruction, path availability model, the secondary routing table.

1 Introduction

Mobile ad hoc networks (MANETs) are self-organized multi-hop wireless networks without requiring any existing infrastructure. Minimal configuration, absence of infrastructure and quick deployment make MANETs convenient for emergency operations. Since no infrastructure exists and node mobility and unpredictable topological changes may cause frequent link failure, it is a great challenge to design an effective and adaptive routing protocol[1][2].

To improve the low packet delivery ratio of DSDV, we propose a Double-Links scheme for DSDV(DLDSDV). DLDSDV needs not any new message exchange for invalid route reconstruction. It maintains and updates a secondary routing table, which contains secondary routes for all available destinations. When a node detects a link is broken, through the secondary routing table, the node can find the neighbor quickly which has a valid route to the destination. Simulation results show that DLDSDV improves the packet delivery ratio, has almost the same routing overhead and end-to-end delay as DSDV. Therefore, DLDSDV can adapt to frequent topology changes quickly in MANETs.

* This research is supported by NSF(Grant No. 61103010, 61103190, 60803100), 863(Grant No. 2012CB933500) and 973(Grant No. 2012AA011001, 2012AA011902).

The remainder of this paper is organized as follows: Section 2 provides the previous work done in this area. The DLDSDV protocol is described in Section 3. In Section 4, the performance analysis of the proposed scheme is presented. Section 5 presents the simulation results. Finally, the paper is concluded in Section 6.

2 Related Works

The efficient DSDV (Eff-DSDV) protocol[2] creates the temporary link by sending one-hop ROUTE-REQUEST and ROUTE-ACK messages. The node upon finding the broken link broadcasts a one-hop ROUTE-REQUEST packet to all its neighbors. In turn, the neighbors returns the ROUTE-ACK if it has a valid route to the destination. Each entry in the routing table has an additional entry for route update time. This update time is used in choosing the latest update time in case of receiving multiple ROUTE-ACK with the same number of minimum hops.

The improved DSDV(I-DSDV) protocol[3] uses a novel message exchange scheme for its invalid route reconstruction. When route becomes invalid due to link breakage, the node that detects the link breakage tries to create a new loop-free route through message exchange with its neighbors. When route reconstruction in one-hop area is not accomplished, the area of message exchange for invalid route reconstruction is enlarged gradually on demand.

A number of protocol optimizations are also proposed to improve the low packet delivery ratio of DSDV in DSDV-MC[4][5]. Recently, many works is done to compare routing protocols for Ad hoc networks in terms of packet delivery ratio, end to end delay and routing overhead in different environment such as varying number of nodes, speed and pause time [6–8].

3 The Double-Links Scheme for DSDV Protocol

In this section, we analyze the problem of invalid route reconstruction which causes low packet delivery in DSDV and its thinking of problem solving. Then, we introduce our proposed strategies DLDSDV with an example.

In DSDV the low packet delivery is due to the fact that it uses stale routes in case of broken links. When a link to the next hop is broken, any route through the next hop is immediately assigned an infinite metric and an updated sequence number[2][3]. As a result, any route through the next hop becomes a stale route.

The existence of stale route does not mean that there is no valid route to the destination. The packets can be sent through other neighbors who have valid routes to the destination[2]. That is to say, when a node detects that a route in its routing table becomes invalid, probably it can create a new route through one of its neighbors which has a valid route to the same destination[3].

In order to find the neighbor which has a valid route to the destination, we propose DLDSDV. In DLDSDV, each node maintains two routing tables. They both list all available destinations, the metric and next hop to each destination

and a sequence number generated by the destination. One is called main routing table, the other is called the secondary routing table. A route in the main routing table is called a main route and a route in the secondary routing table is named a secondary route[4].

In DLDSDV, the main routing table is maintained and updated as it has been done in DSDV. A secondary route should obey the following rules:

- A secondary route is either a valid one or a invalid one and a invalid secondary route has a infinite metric
- A valid secondary route should share the same metric and sequence number as the corresponding main route
- The next hops of a valid secondary route and a main route are different. These ensure that routes in DLDSDV is loop-free
- At the beginning, all secondary routes are invalid
- If a node receive a route update and its metric and sequence number are the same as the main route and its next hop is different from the main route, the corresponding secondary route is updated
- If a main route is replaced with a secondary route, the secondary route is recognized as a invalid one

Given to network topology in Figure 1(a), we can get S 's two routing tables. As table 1 showed, S has three neighbors, they are D, n_1, n_2 respectively and S has two routes to n_3 . The next hop in the main routing table is D and the next hop in the secondary routing table is n_2 .

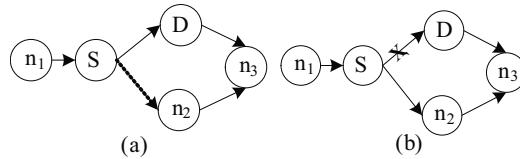


Fig. 1. A simple network topology

When S finds that it is unreachable for D , it does the following approach:

- S checks the main routing table and any main route whose next hop is D should be replaced by the secondary route, then the metric of the secondary route is assigned infinity. So, the metric of the new main route is either the same as the old main route or an infinite one
- The new main route is assigned a new sequence number which is one greater than the old one (call an odd sequence number)
- S advertises all the routes in the main routing table which has an infinite metric or an odd sequence number

Figure 1(b) shows S finds that it is unreachable for D and table 2 shows S 's two routing tables after S finds that it is unreachable for D . As the table 2 depicted, the main route to n_3 is replaced by the secondary route.

Table 1. S 's two routing tables in figure 2(a)

Dest- ination	Main Route			Secondary Route		
	NextHop	Metric	SeqNum	NextHop	Metric	SeqNum
n_1	n_1	1	$seq(n_1)$	-	-	-
D	D	1	$seq(D)$	-	-	-
n_2	n_2	1	$seq(n_2)$	-	-	-
n_3	D	2	$seq(n_3)$	n_2	2	$seq(n_3)$

Table 2. S 's two routing tables in figure 2(b)

Dest- ination	Main Route			Secondary Route		
	NextHop	Metric	SeqNum	NextHop	Metric	SeqNum
n_1	n_1	1	$seq(n_1)$	-	-	-
D	-	-	-	-	-	-
n_2	n_2	1	$seq(n_2)$	-	-	-
n_3	n_2	2	$seq(n_3) + 1$	-	-	-

Each node that hears a routing packet from S which includes a route update to D with an infinite metric, will do the following approach:

- If the node has a main route to D and its next hop is S , the main route is replaced by the secondary route, then the metric of the secondary route is assigned infinity. The metric of the new main route is either the same as the old main route or an infinite one
- The new main route is assigned a new sequence number which is one greater than the old one (call an odd sequence number)
- If the node has a main route to D and its next hop is not S , a sequence number one greater than the old one is assigned(call an odd sequence number)
- If the node has a secondary route to D and its next hop is S , the secondary route revised into a invalid one
- The node advertises all the routes in the main routing table which has an infinite metric or an odd sequence number

In DSDV, incremental update packets are broadcasted between the full dumps for partial changes of the routing table, such as a substantial route change or receiving a new sequence number[1]. According to the description above, in DLDSDV, the node S or the intermediate nodes which receive new sequence numbers will cause an incremental routing update and all the routes in the main routing table which has an infinite metric or an odd sequence number will be broadcasted. Therefore, our proposed scheme does not increase routing overhead.

4 Performance Analysis

In this section, first we introduce a path availability model for MANETs and calculate the path holding probability of the DSDV. Then, we extend the path availability model for MANETs and give formulas of path holding probability of DLDSDV which maintains two routes to all destinations.

4.1 Path Holding Probability Analysis of the Original DSDV

The mobility profile of a given node S moving according to a random ad-hoc mobility model is specified based on three parameters: $\lambda_i, \mu_i, \delta_i$. These parameters and assumptions are defined in [9]. Given two nodes $m < \lambda_m, \mu_m, \delta_m >$ and $n < \lambda_n, \mu_n, \delta_n >$, the probability density function of the equivalent random mobility vector of node m with respect to node n can be written as[10]:

$$f_{z,\theta}(z, \theta) \approx \begin{cases} \frac{1}{2\pi} \frac{z}{\alpha_{m,n}} e^{\frac{-z^2}{2\alpha_{m,n}}} & z \geq 0, 0 \leq \theta \leq 2\pi \\ 0 & \text{others} \end{cases} \quad (1)$$

$$\alpha_{m,n} = \frac{2t(\delta_m^2 + \mu_m^2)}{\lambda_m} + \frac{2t(\delta_n^2 + \mu_n^2)}{\lambda_n} \quad (2)$$

where z is the magnitude of the equivalent random mobility vector, θ is the direction of the vector. z is approximately Raleigh distributed and its parameter is $\alpha_{m,n}$, θ is uniformly distributed over $(0, 2\pi)$. Then we derive the path holding probability of single link between m and n as following[10]:

$$q_{m,n}(d, t) = \int_0^\pi \frac{1 - e^{\frac{-(\sqrt{R_t^2 - d^2 \sin^2 \theta} - d \cos \theta)^2}{\alpha_{m,n}}}}{\pi} d\theta \quad (3)$$

where R_t is wireless transmission radius, d is distance between m and n . For simplicity and without loss of generality we employ mean distance d , to represent the distance of any linked node pair.

Given a transmission pair S and D in DSDV, the total hops between S and D is $n+1$, routes can be expressed as following:

$$\{S, X_1, X_2, \dots, X_n, D\}. \quad (4)$$

We can get the path holding probability of the path from S to D as following:

$$H(t) = q_{S,X_1}(d, t) \times q_{X_n,D}(d, t) \times \prod_{i=1}^{n-1} q_{X_i,X_{i+1}}(d, t) \quad (5)$$

4.2 The Path Holding Probability Analysis of DLDSDV

In DLDSDV, each node maintains two routing tables, the main routing table and the secondary routing table. They both list all available destinations, the metric and next hop to each destination and a sequence number generated by the destination node. Based on descriptions of section 4, a valid secondary route should obey the following rules:

- A valid secondary route should share the same metric and sequence number as the corresponding main route
- The next hops of a valid secondary route and a main route are different. These ensure that routes in DLDSDV is loop-free

Given a transmission pair s and D , the total hops between S and D is $n+1$, the main routes and the secondary routes can be expressed as following respectively:

$$\{S, X_1, X_2, \dots, X_n, D\}, \{S, Y_1, Y_2, \dots, Y_n, D\}. \quad (6)$$

where the first hop and the last hop in the main routing table are $\{S, X_1\}$ and $\{X_n, D\}$ respectively and the first hop and the last hop in the secondary routing table are $\{S, Y_1\}$ and $\{Y_n, D\}$ respectively. The i th hop in the main routing table is $\{X_{i-1}, X_i\}$ and $i \in [2, n]$. The i th hop in the secondary routing table is $\{Y_{i-1}, Y_i\}$ and $i \in [2, n]$. Therefore S has two routes to D and these two routes may be either independent or dependent. If $\exists i Y_i = X_i, i \in [1, n]$, we define these two routes as dependent routes. If $\forall i, i \in [1, n], Y_i \neq X_i$, we define these two routes as independent routes. If $\forall i, i \in [1, n], Y_i = X_i$, we define these two routes as the same routes.

After a careful analysis, we find that there are at most four kinds of mapping modes over one hop. Table 3 depicts these kinds of mapping modes in the i th hop and $i \in [2, n]$. The first mapping mode means there are only one transmitter and one receiver and they compose a one-to-one mapping. The second mapping mode means there are two receivers and one transmitter and they compose a one-to-two mapping. The third mapping mode means there are two transmitters and one receiver and they compose a two-to-one mapping. The fourth mapping mode means there are two transmitters and two receivers and they compose two one-to-one mappings. Given to the first hop and the last hop, they are different. There are only one transmitter in the first hop and one transmitter in the last hop. Table 3 gives descriptions of them respectively. Based on the analysis above, we find we can employ a modified binary tree to depict the correlation of them.

Figure 2 gives a modified binary tree when the total hops from S to D are five. As the figure depicted, node ' a'_{ij} ' means the path holding probability in the i th hop, ' i ' means the i th hop and ' j ' is used to choose mapping mode in the i th hop. Later we will give definition and numerical expressions of a_{ij} . For simplicity and without loss of generality we employ mean distance d , to represent the distance of any linked node pair. The weight ' p'_{ij} ' on the link represents the probability that the mapping mode, decided by ' j ', occurs in the i th hop.

- If $i > 0, j \bmod 4 = 0$, it means the first mapping mode occurs in the i th hop. That is to say there are only one transmitter and one receiver. The path holding probability over this hop can be computed as following, where d is mean distance between of any transmission pairs.

$$a_{ij} = q_{x_{i-1}, x_i}(d, t) \quad (7)$$

- If $i > 0, j \bmod 4 = 1$, it means the second mapping mode occurs in the i th hop. That is to say there are one transmitter and two receivers. The path

Table 3. All of mapping modes in the i th hop, in the first hop and in the last hop

The i th hop	$X_i = Y_i$	$X_i \neq Y_i$
$X_{i-1} = Y_{i-1}$	$x_{i-1} \rightarrow x_i$	$x_{i-1} \rightarrow x_i$ $x_{i-1} \rightarrow y_i$
$X_{i-1} \neq Y_{i-1}$	$x_{i-1} \rightarrow x_i$ $y_{i-1} \rightarrow x_i$	$x_{i-1} \rightarrow y_i$
The first hop	$s \rightarrow x_1$	$s \rightarrow x_1$ $s \rightarrow y_1$
The Last hop	$x_n \rightarrow d$	$x_n \rightarrow d$ $y_n \rightarrow d$

holding probability over this hop can be computed as following, where ρ_1 is a coefficient of relativity which represents the probability that broken link is caused by transmitter.

$$a_{ij} = 1 - (1 + \rho_1)(1 - q_{x_{i-1}, x_i}(d, t))(1 - q_{x_{i-1}, y_i}(d, t)) \quad (8)$$

- If $i > 0, j \bmod 4 = 2$, it means the third mapping mode occurs in the i th hop. That is to say there are two transmitters and one receiver. The path holding probability over this hop can be computed as following, where ρ_2 is a coefficient of relativity which represents the probability that broken link is caused by receiver.

$$a_{ij} = 1 - (1 + \rho_2)(1 - q_{x_{i-1}, x_i}(d, t))(1 - q_{x_i, x_i}(d, t)) \quad (9)$$

- If $i > 0, j \bmod 4 = 3$, it means the fourth mapping mode occurs in the i th hop. That is to say there are two transmitters and two receivers. The path holding probability over this hop can be computed as following.

$$a_{ij} = 1 - (1 - q_{x_{i-1}, x_i}(d, t))(1 - q_{x_i, x_i}(d, t)) \quad (10)$$

Based on the modified binary tree, we can compute the path holding probability of a path from S to a leaf node a_{mn} including the probability. We have

$$\begin{aligned}
H_{mn}(t) &= p_{mn} \times a_{mn} \prod_{i=0}^{m-2} p_{(m-1-i)\frac{n}{2^i}} \times a_{(m-1-i)\frac{n}{2^i}} \\
&= p_{mn} \times \prod_{i=0}^{m-2} p_{(m-1-i)\frac{n}{2^i}} \times a_{mn} \times \prod_{i=0}^{m-2} a_{(m-1-i)\frac{n}{2^i}} \\
&= P_{mn} \times A_{mn}
\end{aligned} \quad (11)$$

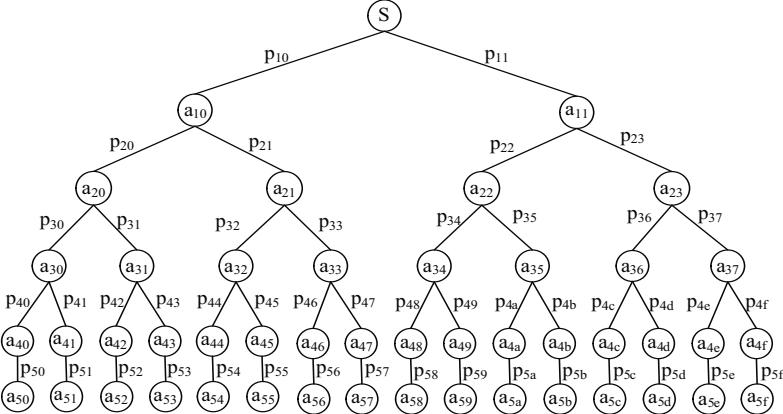


Fig. 2. The modified binary tree

where p_{mn} represents the probability that the mapping mode of a_{mn} occurs in the m th hop. P_{mn} represents the probability that there exists a path from S to a_{mn} , A_{mn} is the path holding probability of a path from S to a leaf node a_{mn} without considering the probability. The computation of P_{mn} is beyond the scope of our paper, people can refer to [9][10]. Then the path holding probability of a whole path from S to D when there exists two dependent routes can be computed as the following equation

$$H(t) = 1 - \prod_{j=0}^{2^m-1} (1 - H_{mj}(t)). \quad (12)$$

5 Simulation Results

We performed simulations using the NS2 simulator[11], its version is 2.29. Simulations were based on the network formed by 100 nodes, distributing in a rectangular (1000m1000m) flat area. The IEEE 802.11 MAC protocol, with a channel bandwidth of 2 Mbps, was used as the MAC layer protocol. The transmission range of each node was approximately 250m. Random-Waypoint was selected as the mobility model. We run our simulations with movement patterns generated for a fixed maximum-speed (20m/s) and 9 different pause-times (s): 0, 50, 100, 150, 200, 250, 300, 350, and 400. The packet size is fixed to 512 Bytes. Each simulation was run for 400 seconds.

There are many improving strategies on DSDV, such as The efficient DSDV (Eff-DSDV) protocol[2] and the improved DSDV(I-DSDV) protocol[3]. In this paper, the performance of DSDV, I-DSDV and DLDSDV is evaluated based on the three performance metrics which are Packet Delivery ratio, End-to-End Delay and the Routing Overhead. Figure 3 shows the routing overhead, packet

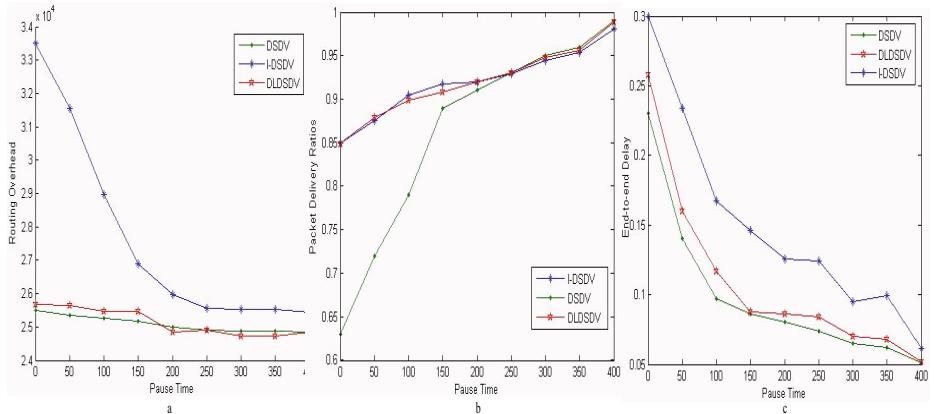


Fig. 3. The comparison of routing overhead, packet delivery ratio and average end-to-end delay

delivery ratios and end-to-end delay of DSDV, I-DSDV and DLDSDV as a function of node pause time. Figure 3 reveals that DLDSDV can deliver more than 85% data packet regardless to mobility rate and it can almost achieve the same packet delivery ratios as I-DSDV can do. DSDV performs poorly as the pause time decreased.

It can be seen from the Figure 3 that, contrast to I-DSDV, DLDSDV exhibits shorter average end-to-end delay all the time regardless to node pause time. When a link is broken, DLDSDV can quickly find the neighbor which has a valid route to the destination according to the secondary routing table, and buffered data packets are transmitted through new route. So DLDSDV exhibits shorter end-to-end delay of data packet compare to I-DSDV. Figure 3 shows, compared to I-DSDV, DLDSDV needs fewer routing packets regardless to node pause time. It can be seen from the graph that DSDV has almost constant overhead regardless of node pause time. I-DSDV needs more routing packets than DSDV and DLDSDV at all time, and their difference becomes smaller as pause time increases.

6 Conclusions

In this paper we propose Double-Links scheme for DSDV which achieves route reconstruction without any new message exchange. It maintains and updates a secondary routing table which contains secondary routes for all available destinations. We extend the path availability model for MANETs and based on the modified binary tree, derive an expression of path holding probability of DLDSDV. Simulation results show that compared with DSDV, DLDSDV can adapt more quickly to frequent topology changes in MANETs. The DLDSDV has less end-to-end delay and routing overhead than I-DSDV.

References

1. Perkins, C.E., Bhagwat, P.: Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers. In: Proc. ACM SIGCOMM 1994, London, UK, pp. 234–244 (October 1994)
2. Khan, K.U.R., Reddy, A.V., Zaman, R.U., Reddy, K.A., Harsha, T.S.: An Efficient DSDV Routing Protocol for WirelessMobile Ad Hoc Networks and its PerformanceComparison. In: The Second UKSIM European Symposium on Computer Modeling and Simulation, India, pp. 506–511 (2008)
3. Liu, T., Liu, K.: Improvement on DSDV in Mobile Ad Hoc Networks. In: International Conference on Wireless Communications, Networking and Mobile Computing, China, pp. 1637–1640 (2007)
4. Fotis Diamantopoulos, A.A.: Economides: A performance study of DSDV-based CLUSTERPOW and DSDV routing algorithms for sensor network applications. In: Proc. 1st International Symposium on Wireless Pervasive Computing, pp. 1–6 (January 2006)
5. Lee, U., Midkiff, S.F., Park, J.S.: A Proactive Routing Protocol for Multi-Channel. In: Proc. Information Technology: Coding and Computing (ITCC 2005), Las Vegas, Nevada, vol. 2, pp. 710–771 (April 2005)
6. Kumar, R., Joshi, V., Dhir, V.: Performance Comparison of Routing Protocols in Mobile Ad Hoc Networks. International Journal of Engineering Science and Technology 2(8), 3494–3502 (2010)
7. Tyagi, S.S., Chauhan, R.K.: Performance Analysis of Proactive and Reactive Routing Protocols for Ad hoc Networks. International Journal of Computer Applications 1(14), 0975–8887 (2010)
8. Rahman, A.H.A., Zukarnain, Z.A.: Performance Comparison of AODV, DSDV and I-DSDV Routing Protocols in Mobile Ad Hoc Networks. European Journal of Scientific Research 31(4), 566–576 (2009)
9. Wan, P.-J., Yi, C.-W.: Asymptotic critical transmission radius and critical neighbor number for k-connectivity in wireless ad hoc networks. In: Proceedings of ACM MobiHoc 2004, p. 108. ACM Press (2004)
10. Liu, J., Guo, W., Xiao, B.-L., Huang, F.: Path Holding Probability Based Ad Hoc On-Demand Routing Protocol. Journal of Software 18(3), 693–701 (2007)
11. Fall, K., Varadhan, K.: The ns Manual, University of Southern California, Information Sciences Institute, ISI (2002),
<http://www.isi.edu/nsnam/ns/nsdocumentation.html>

Time Frequency Analysis of Underwater Carrier-Free Pulse

Yunlu Ni and Hang Chen

School of Marine Engineering, Northwestern Polytechnical University, China

Abstract. Currently realizable carrier-free millisecond or microsecond pulse for novel sonar system has advantages: attaining more target information, restraining fluctuation of reverberation envelop efficiently in short-range detection and achieving accurate estimation. Detection of such short pulse is one of significant tasks for high-resolution sonar system. However, Fourier Transform spectrum of short pulse locates widely and transiently in frequency domain without detailed information, which is difficult to separate short pulse from noise. Short Time Fourier Transform, Wavelet Transform and Hilbert-Huang Transform were employed to analyze short pulse to obtain different time-frequency characteristics, so as to seek for detection methods. Time-frequency characteristics for three time-frequency method are largely identical but with minor differences. A tri-channel detector is assumed for short pulse based on differences of three time-frequency characteristics.

Keywords: Carrier-free Pulse, Time-frequency Analysis, Characteristics.

1 Introduction

Carrier-free short pulse as a transient signal has a relatively wide bandwidth, with considerable information, anti-jamming capability, good stealth and anti-stealth effect, etc. It is widely applied in detection, imaging, accurate orientation and target identification. Brought into Sonar system, its unique advantages are: 1) detecting invisible or quiet target because of broadband characteristics of both low-frequency and high-frequency; while stealthy design of the target is always effective on certain frequency band, as long as Sonar system has a sufficiently wide signal bandwidth invisible target can be detected; 2) high range resolution, distinguishing between the target scattering points; 3) a good ability of identification, separation of the response of different target in different areas, highlighting the target characteristics for identification; 4) ultra-short-range detection capability; 5) to reduce the reverberation fluctuation [1]; 6) anti-jamming performance.

Carrier-free short pulse as Ultra-Wideband Radar signal has mature detection theory in time domain, such as short-term correlation detection [2], segmented correlation detection [3], peak energy detection [4], frequency detection[5], compressed sensing detection[6], etc.. More complex detection method, such as a dual-channel Power-law detector constituting the peak energy detection based on energy detection and Power-Law detection method based on higher order spectrum, applies only to

non-Gaussian underwater acoustic transient signals [7]. In this paper carrier-free short pulse with Gaussian envelope of the signal energy is detected only in the case of high signal-noise-ratio, so higher-order statistics can not distinguish between Gaussian noise and such short pulse.

Short pulse with short time duration has the characteristics of transient signals. Its time-frequency characteristics are fundamental for target detection and identification which has important applications in the detection of quiet target. For future detection, three time-domain frequency methods are employed for obtaining spectrum characteristics to compose a tri-channel detector.

2 Time-Frequency Analysis Principles

Instantaneous wide bandwidth in frequency domain is an important feature of carrier-free short pulse. Traditional Fourier Transform(FT) is a kind of global transformation without detailed frequency information. Three time-frequency methods are employed to analyze short pulse and obtain its time-frequency characteristics, which provides theoretical basis for detection.

2.1 Short Time Fourier Transform

Short Time Fourier Transform (STFT) can be regarded as local spectrum near analysis time, STFT of discrete time signal $x(n)$ may then be defined as[8]

$$STFT = \sum_{m=-\infty}^{\infty} w(n-m)x(m)e^{-j\omega_k m} \quad (1)$$

Equation (1) shows that $w(n)$, the window, selectively determines the portion of $x(n)$ which is being analyzed.

In STFT, discrete signal $x(n)$ first multiplies by analysis window central at time t before Fourier transform. As window function's time shift and frequency shift modulation, it's equivalent to obtain a slice of signal in the vicinity of analysis time point, that is, the local spectrum. Slice at any time t is the local spectrum of the signal at the moment. More important is that once a window has been chosen for the STFT, the time-frequency resolution is fixed over the entire time-frequency plane.

The spectrum of STFT is defined as square of modulus value:

$$SPEC(t, f) = |STFT(t, f)|^2 \quad (2)$$

2.2 Wavelet Transform

In a Wavelet Transform (WT), the notion of scale is introduced as an alternative to frequency, leading to a so-called time-scale representation and having multi-resolution characteristics, which features are suitable for the detection of transient signals. WT uses short windows at high frequencies and long windows at low frequencies, as is in

the spirit of so-called “constant-Q” or constant relative bandwidth frequency analysis. It could effectively focus on the instantaneous part of the signal and overcome the resolution limitation of the STFT. WT is defined as follows[9]:

$$WT_s(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \varphi^*(\frac{t-b}{a}) dt \quad (3)$$

The scalogram is defined as the square of WT modulus values:

$$SCAL(a,b) = |WT(a,b)|^2 \quad (4)$$

2.3 Hilbert-Huang Transform

Researches have shown that prementioned STFT and WT are time-frequency analysis based on FT, and inevitably influenced by defects of FT, such as fake frequency and redundant signal component. Empirical Mode Decomposition (EMD), recently been pioneered by N.E. Huang et al[10], breaks the limitations of FT. Especially for transiently and locally signal, Rilling[11] modified EMD to local EMD.

Given a signal $x(t)$, the effective algorithm of local EMD can be summarized as follows:

- 1) identify all extrema of $x(t)$
- 2) interpolate between minima (resp. maxima), ending up with some envelope $e_{\min}(t)$ (resp. $e_{\max}(t)$)
- 3) compute the mean $m(t) = (e_{\min}(t) + e_{\max}(t))/2$
- 4) extract the detail $d(t) = x(t) - w(t)m(t)$
- 5) iterate on the residual $m(t)$

Through sifting process, Intrinsic Mode Function(IMF) are obtained for spectrum as

$$X(t) = \sum_{j=1}^n a_j(t) \exp(i \int \omega_j(t) dt) \quad (5)$$

3 Target Echo

In this paper, the carrier-free narrow pulses is generalized Gaussian signal, widely used in ultra-wideband signals in the radar and now studied as underwater acoustic emission signal. The waveform[12] is

$$s(t) = \frac{E}{1-\alpha} (\exp(-4\pi(t-t_m)^2 / \Delta t^2) - \alpha \exp(-4\pi\alpha^2(t-t_m)^2 / \Delta t^2)) \quad (6)$$

Where E is peak value of energy, α is Gaussian spread parameters (s^{-1}), Δt is nominal duration, t_m is the time of peak energy.

With $E=1$, $\alpha=1$, $\Delta t=0.1ms$, the signal has instantaneous 3dB bandwidth of 18.8kHz and central frequency is 17.6kHz, in accordance with the definition of ultra-wideband signal, whose relative bandwidth is much larger than 0.25. This short pulse signal is broadband and can not neglect the transmission loss. Considering transmission loss in underwater channel and target effect, pulse width becomes wider provided one highlight point. The target echo is seen in Fig1.

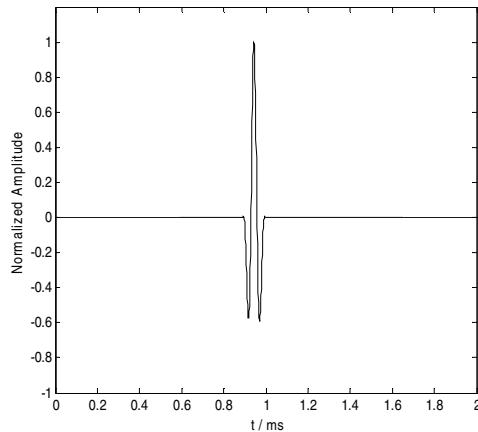


Fig. 1. Simulated target echo

4 Time-Frequency Analysis Result

The spectrum distributes widely at low frequencies with Gaussian envelop. Relative bandwidth is larger than 1. After STFT, the spectrum is presented in Fig2. The time-frequency distribution focuses on pulse time, while the energy distribution at different frequency is observable. From the contour line as Fig3, detailed information and peak energy time are obtained. Because STFT has the same resolution at different

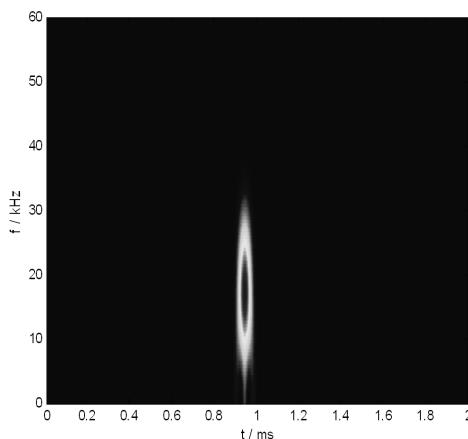


Fig. 2. STFT spectrum

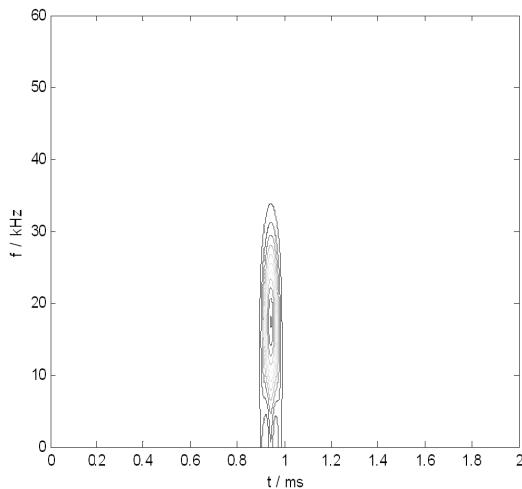


Fig. 3. Contour of STFT spectrum

frequency block, small energy of short pulse can be seen, such as the part of 30kHz-35kHz and 0kHz-5kHz. Energy summation in pulse width in contour can be employed for detection.

After Morlet Wavelet Transform, the scalospectrum is demonstrated in Fig4. Because of the imperfect matching of Morlet wavelet, pulse width extends wider. However, Morlet WT gives prominence to low frequency. For contour and scalogram, time-frequency distribution is waterdrop-shaped, energy focus in low frequency for more time duration. Small energy of some frequency is lost but not important for time-frequency distribution. Different energy distribution in pulse width can be used for such short pulse detection.

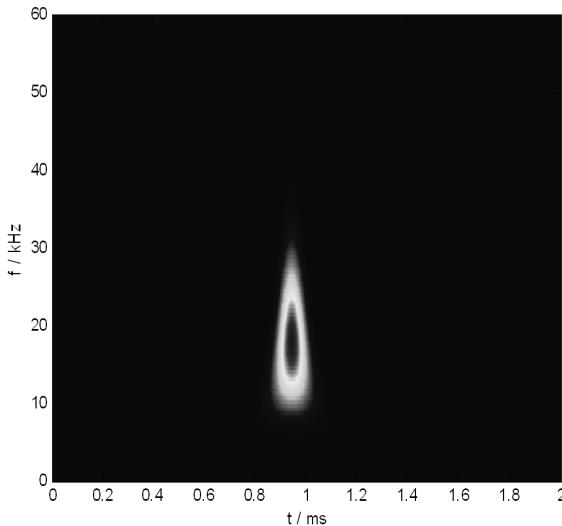


Fig. 4. Scalogram of Morlet WT

After local EMD, HHT spectrum is shown in Fig6. In this part there are three main IMFs presented. Because EMD acts as dyadic filter, peak frequency of third IMF is almost half of that of second IMF. HHT has the advantages of fine resolution, while HHT spectrum is approximately equivalent to contour information. Different IMFs have same low frequency part and extraordinary high frequency part. As this matter, special fine resolution can be used for target recognition.

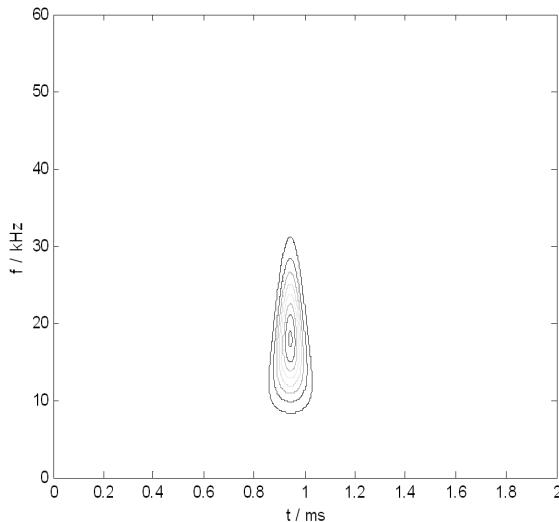


Fig. 5. Contour of Scalogram of Morlet WT

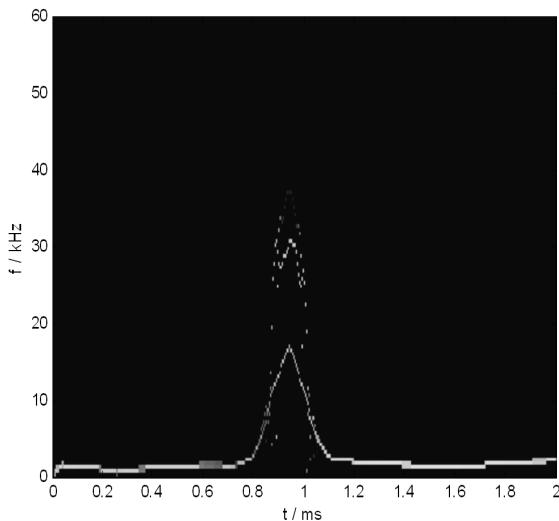


Fig. 6. HHT spectrum

Combining three time-frequency characteristics, a tri-channel detector can be employed for short pulse detection and recognition, with pulse width energy based on STFT, energy distribution in pulse width based on WT and precise recognition based on HHT.

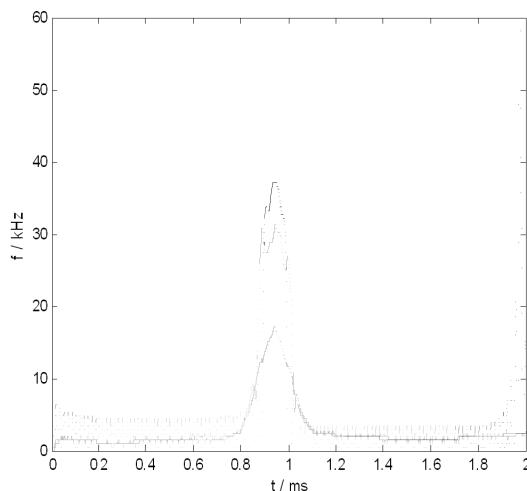


Fig. 7. Contour of HHT spectrum

5 Conclusions

Carrier-free short pulse with transient wide bandwidth has many advantages for sonar system. Relative wideband takes difficulties in short pulse detection. Simple Fourier Transform can not demonstrate the frequency distribution. With time-frequency analysis method, different time-frequency characteristics appear.

For STFT, spectrum energy distributed uniformly and small spectrum energy remains; for WT, scalospectrum demonstrates waterdrop-shaped, with high low frequency energy; for HHT, fine resolution is capable for target recognition. With the characteristics, a tri-channel detector is assumed for short pulse and its performance is put forward to future research.

References

- Chen, H., Chen, Y.S., Yang, H., Ni, Y.L.: A Method using Stochastic Derivative for Analyzing Fluctuation of Reverberation Envelop. *Sci. China Ser. G* 39, 1584–1588 (2009)
- Li, M., Yao, L.: Detection of Transient Signal based on Short-time Correlation. In: Proceedings of National Acoustics Conference, pp. 101–102. China Institute of Acoustics Press (2002)
- Jing, Z.H., Hua, J., Yang, X.N.: A New Method for Detecting Negative SNR UWB Signal. *Communication Countermeasures* 4, 17–21 (2007)

4. Wang, J., Huang, Z.Z., Wang, W.D., Wang, D.J.: A New Peak Detection UWB Impulse Radio Receiver. *Journal of University of Science and Technology of China* 38, 1168–1172 (2008)
5. Zhang, X.Y., Liu, Y.L., Wang, K.: Ultra Wide-Band Channel Estimation and Signal Detection. *Journal of Xi'an Jiao Tong University* 44, 88–91 (2010)
6. Dai, X., Fang, G.Y.: Cramer-Rao Lower Bound of Vital Signal Frequency Detection for Ultra Wideband Radar. *Journal of Electronics and Information Technology* 33, 701–705 (2011)
7. Jun, X., Zhong, L.: Non-Gaussian Underwater Acoustic Transient Signals Detection with Dual-channel Using Power-Law Detectors. *Technical Acoustics* 27, 128–130 (2008)
8. Alllen, B.J., Rabiner, R.L.: A Unified Approach to Short-Time Fourier Analysis and Synthesis. *Proceeding of the IEEE* 65, 1558–1564 (1977)
9. Rioul, O., Vetterli, M.: Wavelets and Signal Processing. *IEEE Signal Processing Magazine* 8, 14–38 (1991)
10. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis. *Proc. R. Soc. Lond. A.* 454, 903–995 (1998)
11. Rilling, G., Flandrin, P., Gonçalvès, P.: On Empirical Mode Decomposition and its Algorithm. In: *Proceedings of the 6th IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP 2003)*, Grado, Italy (2003)
12. Hussain, M.-G.M.: Principles of High-Resolution Radar based of Nonsinusoidal wave—part I: Signal Representation and Pulse Compression. *IEEE Trans. Electromagn. Compat.* 31, 359–367 (1989)

Data Aggregation Scheme Based on Compressed Sensing in Wireless Sensor Network

Guangsong Yang¹, Mingbo Xiao², and Shuqin Zhang²

¹ School of Information Technology, Jimei University, Xiamen, 361021, China

² School of Communication Engineering, Hangzhou Dianzi University,
Hangzhou, 310018, China
gsyang@jmu.edu.cn, mingbo@hdu.edu.cn

Abstract. Compressive sensing (CS) is a novel approach to achieve much lower sampling rate for sparse signals. In order to reduce the number of data transmissions and save more energy, we apply CS theory to gather and reconstruct the sparse signals in energy-constrained large-scale wireless sensor network(WSN).Instead of sending full pair-wise measurement data to a sink, each sensor transmits only a small number of compressive measurements. The processes of CS aggregation in WSN are given. The relationship between Observations and reconstruct MSE are also discussed. Simulation result shows that our scheme can recovery the unknown data accurately as well as reduce global scale cost.

Keywords: Compressed Sensing, Wireless Sensor Network, Aggregation.

1 Introduction

Compressive sensing [1][2]is a collection of recently proposed sampling methods in information theory which deals with estimating an unknown signal with fewer measurements than the Nyquist sampling theorem dictates. Compressed sensing (CS) theory has recently become widely popular to improve system efficiency in the field of image processing, geophysics, medical imaging, computer science, as well as in Wireless Sensor Network (WSN). The two key ideas of CS are sparsity and incoherence. The former depends on single itself, the latter depends on both single and sensing environment.

It is well known that proper data aggregation techniques may reduce the amount of data transmission load carried by a WSN and may hence improve its performance in every aspect. However, conventional aggregation techniques just extract some statistical characteristics from sensing data and loss some features [3], other technology such as Slepian-Wolf coding [4], need using non-cooperative data compress, without prior knowledge of the data correlation structure could render it impossible to perform the coding operations. Collaborative in-network compression makes it possible to discover the data correlation structure through information exchange [5], the resulting high computation and communication load may potentially offset the benefit of this aggregation technique.

CS has been envisioned as a useful technique to improve the performance of WSN [6, 7]. It can use for single processing, signal detection [8], channel estimation [9, 10], etc. Current work for aggregation mainly considers single-hop aggregation [11] and data distribution [12]. In this paper, we consider the application of a new decentralized compression technology known as compressed sensing (CS), to in-network data aggregation.

The remaining of our paper is organized as follows. In Section 2, we review the basic CS concepts that are relevant to our problem. We then describe the work process of CS data aggregation for WSN in section 3. In Section 4, we show the simulation results. Finally, we conclude the paper in Section 5.

2 Compressed Sensing Basis

CS theory asserts that one can recover certain signals from far fewer samples than what have been acquired from the sensors, if those signals can be sparsely represented in a proper basis [13]. It also can be used in a WSN scenario with n nodes, each node acquiring a sample (e.g., temperature) x_i . Our goal is to collect the vector $X = [x_1, x_2 \cdots x_n]^T$ at the sink. We say x has an m -sparse representation if there exists a proper basis $\Psi = [\psi_1, \psi_2 \cdots \psi_m]^T$, s.t. $X = \sum_{i=1}^m S_i \psi_i$ and $m \ll n$. Now the CS theory suggests that, under certain conditions, instead of collecting X , we may collect $Y = \Phi X$, where $\Phi = \{\phi_{j,i}\}$ is a $k \times n$ “sensing” matrix whose entries are i.i.d. zero-mean random variables with variance $1/k$. Consequently, we can recover X from Y by solving the convex optimization problem

$$\min \|S\|_1 = \min \sum_i |s_i| \text{ subject to } Y = \Phi \Psi S \quad (1)$$

and letting $X = \Phi \hat{S}$, with \hat{S} being the optimal solution of (1).

The condition that guarantees the correctness of this recovery is given by $k > C \times m \times \log n$, where C is some small constant. The sink needs to collect only $m \ll n$ samples to reconstruct the sensory data represented by the n samples.

According to the description above, Φ and Ψ are two keys to applying CS in WSN. Using pseudo-random number generators to produce the entries of Φ , we can meet the i.i.d. criterion while avoiding actually transmitting Φ by seeding the generators using publicly known numbers. For example, if we associate a specific generator with a node i , the i -th column of Φ , ϕ_i can be generated anywhere with consistent output. Although the Ψ that yields the sparsest representation of X may not be known, wavelets are in general considered as a good candidate for Ψ .

3 Aggregation in WSN Based on CS

In WSN, data generated by different sensors can be jointly processed while being forwarded toward the sink. Data aggregation is the simplest type of in-network

processing which combines data from different sources or nodes into a single entity. In sensor networks, the data gathered by spatially close sensors are usually correlated, so the node information is compressible. We can use CS technology to eliminate data redundancy. The algorithm process is as follows.

Step 1. Generation of transformation matrix Ψ

$$\Theta_{N \times 1} = [\theta_1 \quad \theta_2 \cdots \theta_k \quad 0 \cdots 0]_{N \times 1}^T \quad (2)$$

From the analysis in section 2 we know, Ψ is used for making compressible single sparse. We use wavelet basis compressed algorithm or 5/3 wavelet lifting algorithm.. Assume there is a perfect optimal basis Ψ , which can make original data $X_{N \times 1}$ sparse, the compressed data is $\Theta_{N \times 1}$ with k nonzero values from beginning.

Step 2. Generation of observation matrix Φ

The $N(0,1)$ independent identically distributed observation matrix meet the requirement of CS theory [14], the element in matrix almost not correlated with sparse element, the observation matrix Φ is bellow:

$$\Phi = \begin{bmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,N} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,N} \\ \vdots & \vdots & & \vdots \\ \phi_{M,1} & \phi_{M,2} & & \phi_{M,N} \end{bmatrix}_{M \times N} \quad (3)$$

The sink needs to store all the n seeds, such that it can generate Φ in order to process the compressed data. In N nodes, the nodes which store nonzero value (the former k nodes), generates M (where, $M = k \log N$) elements which are $N(0,1)$ distribution and mutually independent $\{\phi_{i,j}\}_{i=1}^M$,

Step 3. After the Step 1 and Step 2, the selected node generates locally values, The k column includes M elements $\{Node_i\}_{i=1}^k$ which generate by node k , and multiply $\{\phi_{i,j}\}_{i=1}^M$ by itself data θ_i , i.e. $\Phi * \Theta$, and send it to sink.

So the data received by SINK is

$$y = [\sum_{i=1}^k \theta_i * \phi_{1,i} \quad \sum_{i=1}^k \theta_i * \phi_{2,i} \quad \cdots \quad \sum_{i=1}^k \theta_i * \phi_{M,i}]^T \quad (4)$$

Step 4. Reconstruct at sink

The sparse vector θ can be accurately recovered from y using the reconstruction techniques, one of those is Orthogonal Matching Pursuit (OMP), which most potent is fast implementation and easy to realize, especially apply to energy limited WSN. When SINK received the M values, it can use OMP algorithm to rebuild \hat{y} , and then compute the Mean Squared Error (MSE).

4 Simulation and Evaluation

We set $N = 256$, $k = 60$ ($M = k \log N$, i.e. $M = 144$). Firstly, we transform the original data in figure 1 to k -sparse matrix by transform matrix Ψ , so the N dimensions X will be compressed to k -sparse single Θ in figure 1.

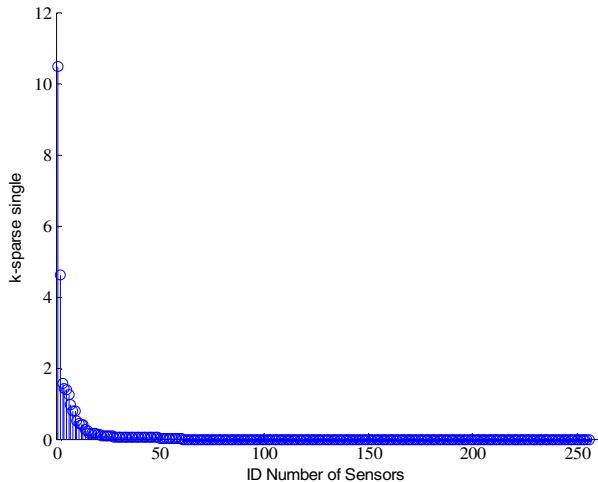


Fig. 1. Sparse coefficient Θ

The data received by SINK is show in figure 2.

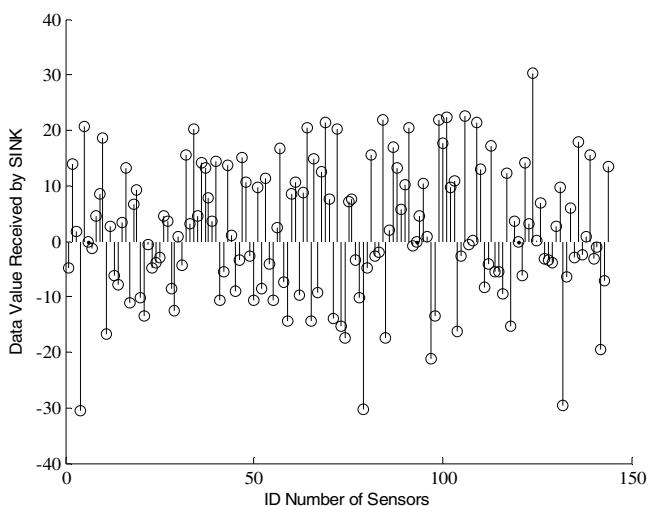


Fig. 2. Received data by SINK

Finally, when SINK received y , recovery the data by OMP algorithm, we can estimate the $\hat{\Theta}$ from $\hat{\Theta}$, figure 3 is the comparison between original data and reconstructed data.

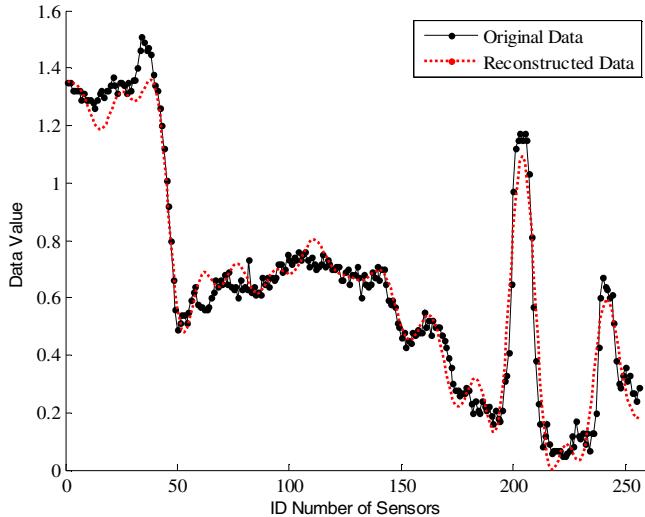


Fig. 3. Comparison between original data and reconstructed data

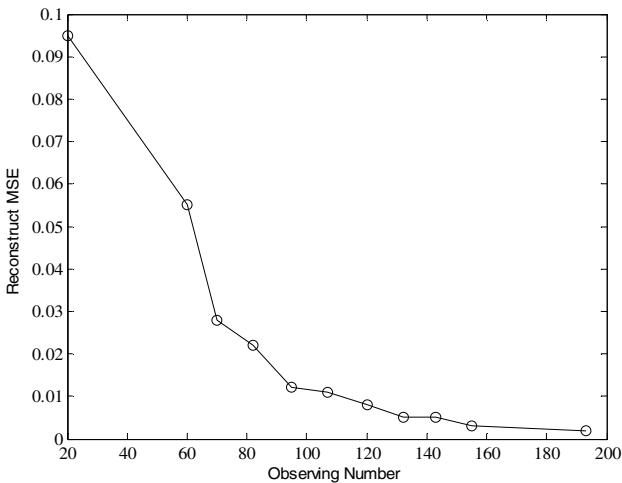


Fig. 4. Observations and reconstruct MSE

We select different observations M , and get different reconstructed precision. The more the observations, the more accurate of data are reconstructed. Figure 4 is the relationship between observations and reconstruction MSE. From figure 4 we can see, the MSE is below 0.02 when observations larger than 80, we can recover the data efficiently.

5 Conclusion

We have described in this paper a novel scheme for overload efficient data gathering in wireless sensor networks based on compressive sensing theory. Based on CS, we design the scheme include design transformation matrix, observation matrix, recovery algorithm in a practical WSN scenario, simulation result show that our scheme can reconstruct the original data at high precision.

Acknowledgments. The work of this paper is supported by Fujian Science Foundation (No. 2011J01356), and Foundation for Young Professors of Jimei University, China (No.2008B002).

References

1. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inform. Theory* 52(4), 1289–1306 (2006)
2. Eldar, Y.C., Kutyniok, G.: *Compressed Sensing: Theory and Applications*. Cambridge University Press (2011)
3. Gao, J., Guibas, L., Milosavljevic, N., Hershberger, J.: Sparse Data Aggregation in Sensor Networks. In: Proc. of the 4th ACM IPSN (2007)
4. Tu, Z., Li, J., Blum, R.: An efficient turbo-binning approach for the Slepian-Wolf source coding problem. *Eurasip Jour. on Applied Signal Processing - Special Issue on Turbo Processing* (2003)
5. Zhang, P., Peng, C., Liu, M.: The application of compressed sensing in wireless sensor network. In: 2009 International Conference on Wireless Communications and Signal Processing, vol. (5) (2009)
6. Haupt, J., Bajwa, W., Rabbat, M., Nowak, R.: Compressed Sensing for Networked Data. *IEEE Signal Processing Mag.* 25(3) (2008)
7. Bajwa, W., Haupt, J., Sayeed, A., Nowak, R.: Compressive wireless sensing. In: Proceedings of the International Conference on Information Processing in Sensor Networks, Washington D.C., USA, pp. 134–142 (2006)
8. Li, C., Sun, T., Kelly, K.F., Zhang, Y.: A compressive sensing and unmixing scheme for hyperspectral data processing. *IEEE Trans. Image Process.* 21(3), 1200–1210 (2012)
9. Taubock, G., Hlawatsch, F.: A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots. In: Proceedings of the IEEE International conference on Acoustics, Speech, and Signal Processing, Washington D.C., USA, pp. 2885–2888 (2008)
10. Bajwa, W.U., Jarvis Haupt, G.R., Nowak, R.: Compressed channel sensing. In: Proceedings of Conference on Information Sciences and Systems, Washington D.C., USA, pp. 5–10 (2008)
11. Rabbat, M., Haupt, J., Singh, A., Nowak, R.: Decentralized Compression and Predistribution via Randomized Gossiping. In: Proc. of the 3th ACM IPSN (2006)
12. Bajwa, W., Haupt, J., Sayeed, A., Nowak, R.: Compressive Wireless Sensing. In: Proc. of the 3th ACM IPSN (2006)
13. Candès, E., Wakin, M.: An Introduction To Compressive Sampling. *IEEE Signal Processing Mag.* 25(3) (2008)
14. Chen, S.B., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 33–61 (1998)

Study of Dynamic Traffic Flow Network Model Based on LWR Model

Wang Jingbo¹, Qian Li¹, and Wang Jingtao²

¹ Hebei United University, Tangshan Hebei, China

² Tangshan Branch Tiandi Science&Technol Co., Ltd., China

jingbow@163.com

Abstract. Modern intelligent transportation requires real-time travel time of dynamic traffic network. Because the traditional model is too complex, it is difficult to provide real-time travel time. A kind of dynamic traffic network analysis model to calculate real-time is given in this paper. The travel time for reference can be computed. Firstly, in the model LWR is used to construct the continuity equation of initial value - boundary condition, then high order Runge-Kutta method is used to calculate of flow, density and speed on the section, for the further travel time required by the vehicle's distance is computed, the total travel time can be obtained by the sum each travel time. Finally, the proposed dynamic traffic network model were simulated for a small network of freeway traffic flows. Simulation results show that the model developed in this paper can accelerate solve the travel time of dynamic network traffic flow.

Keywords: LWR Model, Dynamic Network Model, Travel Time.

1 Introduction

Travel time estimation and forecast information are data of traffic prediction model and the level of road service performance needed. The required travel time of developing intelligent transportation systems becomes from static to dynamic link travel time, or even real time travel time. Especially, it is particularly important for the correctness of real-time link travel time prediction in traffic assignment and dynamic route guidance system.. Otherwise, assignment or guidance error will make the passers-by lost confidence on system then lose the meaning of intelligent transportation system. Addition, in the traffic network management strategy use, the system is unable to select the best execution of a scheme or supporting traffic control measures if there is no dynamic or real-time travel time prediction.

By the related literature of the dynamic traffic flow [1-10], we can know that when a model construction is more complex, accordingly, simulation time is longer to provide real-time traffic flow control information. It is difficult to provide real-time traffic flow control information. In the absence of efficient computational model, this paper decides to use the familiar LWR traffic flow model as the initial architecture of

the developing dynamic analysis model, in order to calculate the travel time and provide real-time information.

Wave equation is used to describe wave propagation phenomena, generally pushed by conservation. According to different problem types and description of the phenomenon, there will be different conservation laws, such as: the string is usually derived from the momentum conservation law i, while wave is derived from flow conservation law. Lighthill, Whitham and Richards will firstly fluid behavior to describe traffic flow behavior and use wave equation to explain the prevalence for car. Therefore, the wave equation applied to describe the traffic behavior is also known as the LWR model [1, 2].

Travel time in traffic network is not only the main basis of the choice for traveler in traffic distribution, but also is an important indicator measuring traffic flow management model .If a vehicle run from starting point to ending point by n path section, then the total travel time of the path is the totally result of vehicle through each path travel time, as shown in figure 1.



Fig. 1. This shows a figure consisting of travel time evaluation model

The related study of travel time prediction model can be broadly divided into static traffic flow model and dynamic traffic flow model, the main difference of which lies in the ability processing information time variance problem. Any incidents that effect on traffic flow in the dynamic analysis of the network can be called traffic incidents, including many factors affecting the network environment, such as vehicle collision accidents, vehicle breakdowns, path construction, and path surface for example, the vehicle on fire and not in weather conditions.

Abroad conducted in-depth research, on the traffic flow simulation, but they are omitted and not talking about it when the issue relates to the dynamic traffic flow model of vehicle travel time to determine the critical issues, and it is yet to see this information in the domestic. Each car's individual behavior should be described in detail in the dynamic traffic flow model, which behavior in the traffic network is determined by the actual traffic situation. In the model, assuming that each car to seek the shortest path, and travel is known in the simulated time, this is the precondition of simulation model.

2 Methods

In this paper, the dynamic traffic network simulation model is to use LWR model to obtain the average link travel time, and then adjustment model operation mechanism through the daily dynamic, achieve flow pattern daily peak hour commuter trip in tends stable. Therefore, it can be divided into two parts. The first part is the LWR model of link travel time calculator system, the second part is the daily dynamic adjustment model, which details are as follows.

2.1 Path Cost Function

In the past, calculating the path cost, we mostly use formula (1), this type of function to calculate the travel time, in which A_a , B_a respectively is parameters of path a , C_a is travel time spending through the path a, and K_a is path a' capacity, f_a is the flow of path a.

$$C_a(f_a) = A_a + B_a \left(\frac{f_a}{K_a} \right)^4 \quad (1)$$

Because the parameters setting should vary with location varies, and parameter estimation is not easy, so using formula (1) the function to calculate the travel time is too general. This paper uses another computing travel time mechanism to obtain the average travel time of path section. This paper uses the running speed of $V(x, t)$ solving by using LWR model when calculation of travel time as a computational tool for travel time.

Fristly, assuming no entrance or cross in the path and the vehicle not appear or disappear, the number of vehicles will satisfy the flow conservation law. The LWR model as shown in (2).

$$\begin{aligned} \frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} &= 0 \\ q(x, t) &= k(x, t) \bullet u(x, t) \\ u &= f(k) \end{aligned} \quad (2)$$

Then, discretizing LWR model, we can get finite difference equation of LWR model. As shown in (3), k_j^n is the density approximation of spatial point x_j and time point t^n , q^n is the flow approximation of point x_{j+1} , point t^n . And Δx is the space distance of discrete model, Δt is the time distance.

$$k_j^{n+1} = \frac{k_{j+1}^n}{2} - \frac{\Delta t}{\Delta x} \cdot \frac{q_{j+1}^n - q_{j-1}^n}{2} \quad (3)$$

Third step, because the speed is density function, we can combine the relationship of flow-density-velocity and the relationship between speed and density to derive the flow-density relation. And we made on the flow of the path f_a into boundary conditions of LWR model, per unit time path of inflow.

The fourth step, we construct the continuity equation of starting value - boundary condition by using LWR traffic flow model with initial condition and boundary condition. We can obtain at each time point, the space point density approximation and velocity approximation.

The fifth step, we estimate the average travel time the section a taking through the velocity approximation. as shown in (4).

$$Ca = \frac{\sum_m \sum_{l_i^m \in I^m, d_j^m \in D^m} \text{Min}\left(\Delta t, \frac{\text{Min}(\Delta x, \varepsilon)}{v(l_i^m, d_j^m)}\right)}{M} \quad (4)$$

Among, Ca said the average travel time must spend through the path a; m said time m into path a fleet; M total fleet numbers. After finding out the average travel time by road, the total time of path travel, the network route travel time, is obtained. And make the values into the daily dynamic adjustment model, we analysis the changes of the flow and path travel time.

2.2 The Daily Dynamic Adjustment Model

In this paper, the daily dynamic adjustment model is base of the dynamic traffic flow network simulation model. It is consists of two differential equations, the change rate of flow and the change rate of ITS travel time forecasting model, such as (5) shown below.

$$\frac{dh_p(t)}{dt} = -\alpha h_p(t)[c_p(h(t) - \hat{c}_w(t))]^r \quad (5)$$

2.3 Solving the Model

In the solution of the model, the LWR model uses finite difference method to solve, so we focus on solving daily dynamic adjustment model. Due to the daily dynamic adjustment model for first-order differential equations, this paper uses the high order Runge-Kutta method for solving because of the high order Runge-Kutta method with high accuracy, which can greatly reduce the error value. For the convenience, using simple way to express the daily dynamic adjustment model, as shown in (6).

$$\begin{aligned} \frac{dh_p(t)}{dt} &= f(h_p(t), \hat{c}_w(t)) \\ \frac{d\hat{c}_w(t)}{dt} &= g(h_p(t), \hat{c}_w(t)) \end{aligned} \quad (6)$$

Solving the daily dynamic adjustment model by third-order of the Runge-Kutta, we can forecast the next day information based on today. When meeting the stop condition, the algorithm stops, and obtains distribution patterns of stable network traffic flow and the forecasting travel time.

3 Results

Using a simple network as the simulation object, the network is a small highway single lane, which geometric form is as shown in figure 2. One group of origin-destination pair of w = (1, 3), the path 1 is made of section 1 and 4, path 2 is made of

section 2 and section 5, and path 3 is made of section 1,3 and 5. The the length of section , $\{L1,L2,L3,L4,L5\} = \{30,55,25,55,30\}$. Model parameters are $\alpha=0.003, \beta=0.11, \gamma=1$, travel demand is fixed to a constant, $T_w(t)=4800$, start path flow, $\{h1,h2,h3\}=\{1400,1800,1600\}$. ITS Initial prediction of travel time $w(t) = 71.0$. The simulation results are obtained as shown in Table 1 and figure 3 and figure 4.

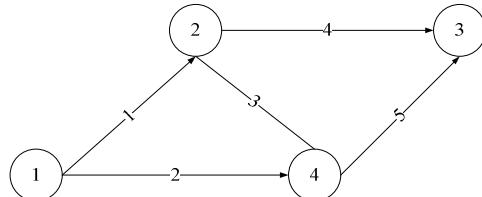


Fig. 2. This shows a figure consisting of an example about network geometry diagram

Table 1. Traffic and travel time starting value and equilibrium value comparison

Path number	Flow		Travel time	
	Initial value	Stable value	Initial value	Stable value
1	1400	1931.6	71.3	71.2
2	1800	1933.7	71.5	71.2
3	1600	934.62	71.7	71.2
4	3000	2866.2	26.20	26.1
5	1800	1933.7	45.07	45.1
7	1600	934.62	19.13	19.0
8	1400	1931.6	45.11	45.1
9	3400	2868.3	26.40	26.1

From table 1 it can be observed that the three path travel time is the same when a steady state is reached (Time Step = 3000), the forecasting travel time also is the same, which represents the traffic flow network to achieve the dynamic equilibrium. In addition, the three path of flow is equal to travel demand in the steady state. Figure 3 is made of the change curve of path 1,2 and 3 travel time and the change curve of ITS predicted travel time. We can find in the simulation of the first half time, curve oscillation is more intense, the back curve of the oscillation is slow, and having the tendency gradually reached dynamic user equilibrium. Figure 4 includes predicted curve of ITS traffic flow and the sum of the actual flow of three paths. It assumes that origin-destination traffic flow is fixed to a constant value, so the traffic flow curve is a horizontal line. From the figure 4, we can find curve oscillation is more intense in first half time, after the oscillation curve slowed, approaches the real traffic flow in the simulation experiment.

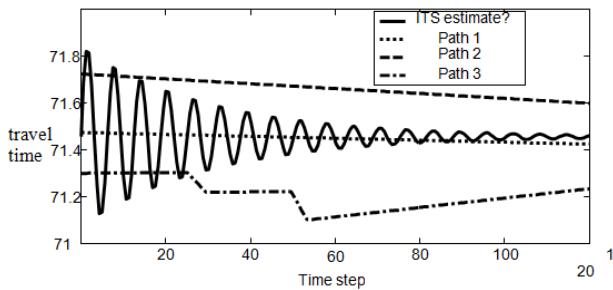


Fig. 3. This shows a figure consisting of the travel time forecasting

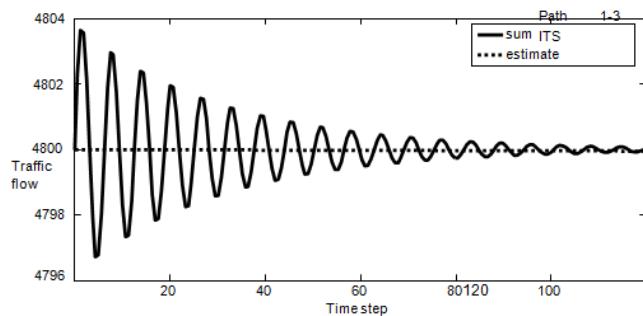


Fig. 4. This shows a figure consisting of the three path flow adjusted

4 Discussion

This paper proposed a new way of estimating travel time; the running speed was calculated as the basis of travel time by the LWR model. In addition, it solved the dynamic traffic network model by high order Runge-Kutta, so that the error reduction. In this paper, it was simulation of a simple network by LWR model, but in the actual network, road environment should be more complex. Therefore, in the future paper, we should be further extended into a model suitable for signalized cross or multi-lane road environment, according with the actual traffic flow network model, in order to reach the purpose of application.

References

1. Adler, J.L.: Investigating the Learning Effects of Route Guidance and Traffic Advisories on Route Choice Behavior. *Transportation Research Part C: Emerging Technologies* 9, 1–14 (2003)
2. Palacharla, P.V., Nelson, P.C.: Application of Fuzzy Logic and Neural Networks for Dynamic Travel Time Estimation. *International Transactions in Operational Research* 6, 145–160 (2004)

3. Smith, B.L., Smith, K.W.: An Investigation into Incident Fraction Forecasting for Fleet Forward. Center for Transportation Studies National ITS Implementation Research Center (UVA-CE-ITS_01-5) (2000)
4. Smith, K.W., Smith, B.L.: Forecasting the Clearance Time of Freeway Accidents. Smart Travel Lab Report 39, 89–112 (2008)
5. Nam, F.M.: An Exploratory Hazard-Based Analysis of Highway Incident Duration. *Transportation Research Part A* 34, 85–102 (2006)
6. Golob, F., Recker, W.W., Leonard, J.D.: An Analysis of the Severity and Incident Duration of Truck Involved Freeway Accident. *Accident Analysis and Prevention* 19, 375–395 (2009)
7. Liping, Z., Juanjuan, W., Honglu, X.: Simulated Prediction of traffic Accident Based on Radial Basis Function Neural Network. *Computer Simulation* 5, 111–114 (2009)
8. Xiaoming, C., Chunfu, S., Da, L., Chunjiao, D.: Capacity Reliability of Signalized Intersections with Mixed Traffic Conditions. *Tsinghua Science and Technology* 14, 333–340 (2009)
9. Juan, L., Chunfu, S., Liya, Y.: Pedestrian detection based on improved Gaussian mixture model. *Journal of Jilin University (Engineering and Technology Edition)* 1, 41–45 (2011)
10. Chunjiao, D., Chunfu, S., Xiaoming, C., Juan, L.: Road network dynamic splitting based on-time-space characteristic of traffic flow. *Journal of Jilin University (Engineering and Technology Edition)* 6, 1528–1532 (2010)

Multiple Sources-Multiple Relays Precoding Algorithm in Cooperative Wireless Networks

Yongli An^{1,2}, Xuefei Gao¹, and Kaiyu Zhu¹

¹ College of Information Engineering, Hebei United University, Tangshan Hebei, China

² Institute of Information Science, Beijing Jiaotong University, Beijing, China

10112059@bjtu.edu.cn

Abstract. Cooperative communication technology can make a wireless network obtain spatial diversity and multiplex gain, specifically in multi-input multi-output cooperative communication system. Cooperative wireless networks have a problem of mutual interference in the scenario of multiple sources-multiple relays transmitting and receiving. Interference alignment technology can solve this problem. We proposed a multiple sources-multiple relays precoding algorithm. We address the issue by giving a system model which transmit four packets through three source users and three relay users. Our simulation results indicate that source node users can transmit more packets at the same time by adopting our proposed algorithm than traditional algorithms.

Keywords: Cooperative Communication, Multi-Sources, Multi-Relays.

1 Introduction

Cooperative communication can improve the diversity gain, coverage and capacity of wireless communication system. It's a combination of MIMO (Multiple Input and Multiple Output) technology and relay communication [1]. Since the mobile device can not be equipped with multiple antennas, so the throughput of MIMO system is limited. Distributed virtual MIMO and relay communication technology can solve this problem.

The traditional relay communication aims to extend the broadcasting distance. The relay node does not only amplify and forward the signal which is received from the source node but also demodulate and decode the signal. It also applies channel encoding technology. According to different patterns in second stage of transmitting signal, the signal processing method can be divided into repetition-based cooperation and coded cooperation [2]. The repetition-based cooperation can also be divided into amplify-and-forward i.e. AF pattern and decoded-and-forward i.e. DF pattern [3]. In AF pattern, the cooperative node just amplifies the signal and transmits it to the base station. The disadvantage of this pattern is that noise of a system is also amplified with the signal and transmitted forward. In DF pattern, the cooperative node decodes the received signal and removes the noise [4]. Cooperative communication has absorbed both advantage of MIMO and relay communication. That means, in a

multiple-users cooperative communication system, source node send date to both relay node and destination node. So in a wireless fading channel scenario, the receivers can get diversity gain and multiplexing gain (through distributed spatial-time coding: each relay node send one row of the precoding matrix).

Cooperative communication makes full use of spatial diversity and multiplex technology [5]. In order to enhance the cooperative diversity gain and multiplexing gain further, related research has proposed distributed space-time encoding theory applying to cooperative communication. [6] Used orthogonal STBCs (Space Time Block Coding) as the distributed STBCs. Each relay node transmitted a different column vector of the STBCs matrix. Regarding to high-capacity cooperative MIMO network, there are a large number of active relay nodes available throughout the network. But it is difficult to design orthogonal STBCs for so many relay nodes and it existed rate loss. [7] Adopted full-rate non-orthogonal STBCs algorithm, while it has not considered cooperative communication technique for its relay nodes.

Currently, most of cognitive networks are based on a network model of single source node and multiple relay nodes, while in this paper we focus on a new network scheme, i.e. multiple source nodes and multiple relay nodes cooperative network. We will address these issues: How does the cooperative network works with multiple source nodes and multiple relay nodes? How can we select relay nodes and how can we do the precoding processing.

2 Cooperative Network Model

Selected researching issue: traditional 3 nodes model had been fully researched as shown in Fig. 1.

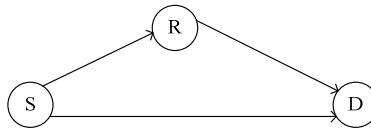


Fig. 1. Traditional 3 nodes model

In multiple relay nodes wireless networks shown as Fig. 2, the selection of relay node is same as single relay communication network. Selection strategy such as: distance-based strategy, maximum SNR strategy are also fully researched. Now much more works focus on multiple relays node even multiple sources node network. So how to choose the relay node from a large amount of active nodes or how to divide source nodes into groups are also need to be studied. These will compose a multiple sources and multiple relays network as shown in Fig. 3.

Multiple relays selection strategy should be adapted to the transmitting protocol. That is to say how the whole network works. That include what kind of coding method or forward method are used. And now a new technique called interference alignment and cancel can also adopt in cooperation. So multiple-sources cooperative network is a good issue worthy to be researched.

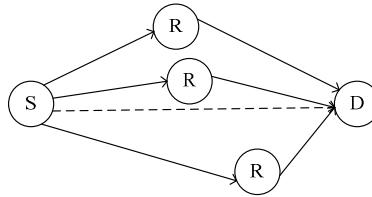


Fig. 2. One source-multiple relays model

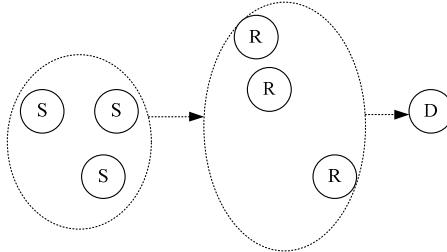


Fig. 3. Multiple sources-multiple relays model

3 The Proposed Network Model

We propose a new network scheme specifically for MIMO wireless network, which composed of multiple source nodes and multiple relay nodes. We express this scheme using 3 source node users and 3 relay node users, cf. Fig. 4.

We will divide wireless network users into two groups, i.e. source node group and relay node group. Compare to traditional single source network, the proposed network model has 3 source node users. These 3 users transmit signal to relay node users at the same time. Through precoding process and linear joint decoding algorithm [8], those 3 relay node users can decode signals received from different source node users without interference.

3.1 Multiple Sources Node Cooperation Algorithm

We proposed a multi-source nodes cooperative encoding and decoding algorithm based on distributed interference alignment. The scheme is that relay node decodes and forwards the signal completely.

Our network model is shown in Fig. 4. S_1, S_2 and S_3 stand for 3 source node users. These 3 users transmit 4 data symbols to relay node users, i.e. R_1, R_2 and R_3 at the same time. x_i is the i th data symbol; H_{ij} is the channel transmitting matrix from i th source node user to j th relay node user; v_i is the precoding vector corresponding to i th data symbol. The linear joint encoding and decoding algorithm focus on the design of precoding vector v_i .

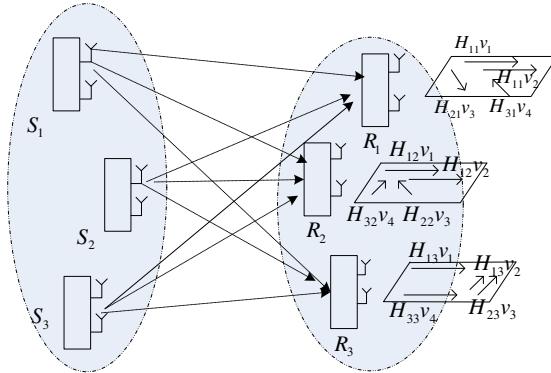


Fig. 4. The proposed multiple sources-multiple relays model

Since the users of cooperative communication wireless networks have two antennas, so the precoding vector v_i is a two-dimensional column vector. Let vector v_i multiply with x_i , then send the result to relay node R_i through wireless channel. Since all these vectors are two-dimensional, they all lie in one-dimensional space.

At the relay node R_1 , we make the vectors $H_{11}v_1$ and $H_{11}v_2$ parallel, while, vectors $H_{21}v_3$ and $H_{31}v_4$ are arbitrary.

Similar with R_1 , we also make the vectors $H_{12}v_1$ and $H_{12}v_2$ parallel, while, vectors $H_{22}v_3$ and $H_{32}v_4$ are arbitrary at the relay node R_2 .

At the relay node R_3 we have the same design that is vectors $H_{13}v_1$ and $H_{33}v_4$ are parallel, vectors $H_{13}v_2$ and $H_{23}v_3$ are parallel, respectively. The precoding vector v_i should satisfy the following equation

$$R_1 : H_{11}v_1 // H_{11}v_2 \quad (1)$$

$$R_2 : H_{12}v_1 // H_{12}v_2 \quad (2)$$

$$R_3 : H_{13}v_1 // H_{33}v_4, H_{13}v_2 // H_{23}v_3 \quad (3)$$

At the relay node R_1 , the receiving equation can be written as

$$y_1 = H_{11}v_1x_1 + H_{11}v_2x_2 + H_{21}v_3x_3 + H_{31}v_4x_4 + n_1 \quad (4)$$

n_i is the AWGN at i th relay node. As long as we find a decoding vector that orthogonal with the vector $H_{11}v_1$ or $H_{11}v_2$, assuming this vector to be Q_1 , i.e. inner product between $H_{11}v_1$ and Q_1 or between $H_{11}v_2$ and Q_1 is zero. This can be written as

$$\langle Q_1, H_{11}v_1 \rangle = 0, \langle Q_1, H_{11}v_2 \rangle = 0 \quad (5)$$

Then we multiply vector Q_1 to both side of (4). We get

$$Q_1 y_1 = Q_1 H_{21} v_3 x_3 + Q_1 H_{31} v_4 x_4 + Q_1 n_1 \quad (6)$$

Similarly, we do the same operation at relay node R_2 . At R_2 , the receiving equation is

$$y_2 = H_{12} v_1 x_1 + H_{12} v_2 x_2 + H_{22} v_3 x_3 + H_{32} v_4 x_4 + n_2 \quad (7)$$

We assume exist a vector P_2 which saitisfy the fallowing condiction

$$\langle P_2, H_{12} v_1 \rangle = 0 \text{ or } \langle P_2, H_{12} v_2 \rangle = 0 \quad (8)$$

Then we multiply vector P_2 to both side of (7). We obtain

$$P_2 y_2 = P_2 H_{22} v_3 x_3 + P_2 H_{32} v_4 x_4 + P_2 n_2 \quad (9)$$

From (6) and (9), when AWGN is ignored, we can achieve the linear decoding of x_3 and x_4 . At relay node R_3 , the receiving equation is

$$y_3 = H_{13} v_1 x_1 + H_{13} v_2 x_2 + H_{23} v_3 x_3 + H_{33} v_4 x_4 + n_3 \quad (10)$$

When we find a vector assuming to be M_3 which is orthogonal to $H_{13} v_1$ or $H_{33} v_4$, then we have

$$M_3 y_3 = M_3 H_{13} v_2 x_2 + M_3 H_{23} v_3 x_3 + M_3 n_3 \quad (11)$$

We substitute x_3 into (11), then we can decode x_2 .

Similarly, when we find a vector assuming to be N_4 which is orthogonal to $H_{13} v_2$ or $H_{23} v_3$, then we have

$$N_4 y_3 = N_4 H_{13} v_1 x_1 + N_4 H_{33} v_4 x_4 + N_4 n_3 \quad (12)$$

We substitute x_4 into (12), then we can decode x_1 .

4 Simulation Result and Analysis

In scenarios of Rayleigh flat fading channel, we simulate the received symbol error rate of relay node using BPSK and QPSK modulation separately. In fact, the simulation process for different relay node users is similar. So we only select one relay node user to simulate. All of the channel fading coefficient matrix $H_{i,j}$ $i, j \in \{1, 2, 3\}$ are generated by Rayleigh channel distribution function, which are all complex matrix of 2×2 order. The simulation result is shown in Fig. 5.

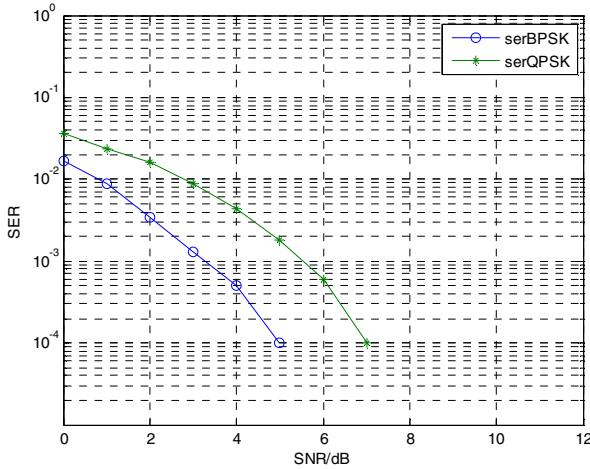


Fig. 5. The curve of symbol error rate at relay node

In Fig. 5, when SNR is larger than 10dB, the theoretical average symbol error rate equals to zero under both BPSK and MPSK modulation pattern. So, this paper proposed a scheme of cooperative MIMO cognitive network composed of multi-source users and multi-relay nodes have high performance of error-resilience. We have two ways to solve the problem of transmission error from source node to relay node. First method is using cyclic redundancy check. But this method increases redundancy and reduce transmission rate of information. The second method is using pre-threshold, i.e., at the relay node, we pre-set a threshold. When the SNR of received signal at the relay node is higher than the pre-threshold, we do the decoding processing. This method is more simple and reliable. From the above simulation result, we can infer that as long as the SNR is greater than 10dB, the relay node can decode without error in theory.

In fact, according to the standardized requirement of future network, such as 3GPP Long Term Evolution LTE network, its downlink rate is 100Mb/s and uplink rate is 50Mb/s. This is the basic requirement. Currently, this target has been exceeded. The bandwidths of LTE networks are 1.4MHz, 5MHz, 10MHz and 20MHz. The famous Shannon theorem shows the smallest SNR requirement for achieving reliable transmission when bandwidth is B and system capacity is C . According to Shannon theorem, when setting the uplink rate to be 50Mb/s and bandwidth to be 10MHz, we can calculate the required minimum SNR is 15dB approximately. Since the SNR of the system can satisfy the requirement of pre-threshold, so the relay node user can decode data without error in theory.

5 Conclusion

In this paper, we proposed a cooperative MIMO cognitive network scheme composed of multiple source node users and multiple relay node users. By dividing users into

two groups, multiple source node users of the proposed scheme can share multiple relay node users simultaneously. They transmit data without interference. Compared with the existing single source node model, adjacent multiple source node users of the proposed scheme can transmit data at the same time. This scheme improved the efficiency of the cooperative MIMO cognitive network. Meanwhile, we have derived the symbol error rate of the system and simulated it at last. Simulation result showed that this scheme has great value and practical significance for LTE even 4G networks.

References

1. Mitola, J.: Cognitive radio for flexible mobile multimedia communications. In: Proc IEEE MoMuC 1999, pp. 3–10 (November 1999)
2. Laneman, N.J., Tse, D.N., Wornell, G.W.: Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transaction on Information Theory* 50(12), 3062–3080 (2004)
3. Stefanov, A., Erkip, E.: Cooperative space-time coding for wireless networks. *IEEE Transactions on Communications* 53(11), 1804–1809 (2005)
4. International Conference on Signal Processing, Beijing, pp. 397–401 (2006)
5. Xiao, Y., Lu, L.Y., Hu, S.H.: Downlink Multi-Beamforming Control for the Smart antennas of TD-SCDMA Base stations. In: Proc. 2004 7th International Conference on Signal Processing (ICSP), Beijing, pp. 451–454 (2004)
6. Xiao, Y., Lu, L.Y., Wang, Y.C.: Space-time spreading in downlink of TD-SCDMA systems. In: Proc. the IEEE 6th Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Communication, vol. 2, pp. 635–638 (2004)
7. Chen, W.J., Xiao, Y., Zhao, Y.: The Algorithm Implement of WCDMA Channel Estimation. In: Proc. 7th International Conference on Signal Processing, pp. 1894–1897 (2004)
8. Gollakota, S., Perli, S.D., Katabi, D.: Interference Alignment and Cancellation. In: Proceedings of the ACM SIGCOMM Conference on Data Communication (SIGCOMM), pp. 159–170 (2009)

Efficient Detect Scheme of Botnet Command and Control Communication

Jinquan Zeng^{1,2}, Weiwen Tang², Caiming Liu³, Jianbin Hu⁴, and Lingxi Peng⁵

¹ School of Computer Science & Engineering, University of Electronic Science and Technology of China, 610054 Chengdu, China

² Sichuan Communication Research Planning & Designing Co., Ltd., 610041 Chengdu, China

³ Laboratory of Intelligent Information Processing and Application, Leshan Normal University, 614004 Leshan, China

⁴ School of Electronics & Information, Nantong University, 226019 Nantong, China

⁵ Department of Computer and Education Software, Guangzhou University, 510006 Guangzhou, China
zengjq@uestc.edu.cn

Abstract. Botnet is an attack network composed of hundreds of millions of compromised computers. Botnet is emerging as the most serious threat against cyber-security and is used to launch Distributed Denial of Service (DDoS) attacks, malware dissemination, phishing, remote control, click fraud, and etc. Although botnet has posed serious security threat on Internet, the research of detecting and preventing botnet is still in its infancy. One effective technique for botnet detection is to identify botnet C&C traffic. In this paper, we present a case study of the IRC-based botnet C&C communication and then present a novel method to detect botnet C&C communications. We develop quantitative ways to assess the C&C communications between the bot and the C&C server; furthermore, we also illustrate the correlation methods within the same botnet's C&C communications to decrease the false positive rate.

Keywords: Bot, Botnet, Computer Security.

1 Introduction

Because botnets pose a significant and growing threat against cyber-security, botnets have been an active research topic in recent years[1-3]. Current research on botnets is focused primarily on botnet principles, botnet detection and botnet samples analysis [4-5]. Botnet principles and botnet samples analysis mainly aim for detection. For understanding botnet principles, many literatures [6-11] presented honeynet-based solutions. However, honeynets are useful to understand botnet principles, but it is difficult to detect botnets. Signature-based detection techniques, such as Anti-Virus tools, are useful to detect only a bot, but cannot destroy whole botnets and cannot detect unknown bots.

In order to detect unknown botnets, anomaly-based detection techniques are used to detect botnets. Gu et al. analyses bot samples and presented a botnet infection

life-cycle dialog model to detect botnets, called as Bothunter; if a bot behavior matches the pre-defined model and then a bot will be reported [12]. Bothunter is based on the pre-defined model to detect botnets, because the pre-defined model cannot cover whole infection life-cycle dialog, such as a computer get infecting by using browsing WEB, so some botnets cannot be detected. Geobl et al. proposed Rishi to detect bot-infected machines within a given network that relies on detection of the communication channel between bot and Command & Control server (C&C server) [13]. However, this approach is quite limited, in that a bot can choose like normal host's IRC nickname and cannot detect non-IRC botnets. The same to [13], their techniques are specific to IRC-based botnets. In addition, their techniques require access to payload content. Thus, it cannot detect encrypted C&C traffic. Xiaocong Yu et al. [14] proposed online botnet detection based on incremental discrete flourier transform.

One effective technique for botnet detection is to identify botnet C&C traffic [4]. Botnet C&C communication is the most important factor in botnets and it is the communication base of botnets. In this paper, we present a case study of IRC-based botnet C&C communication and then present a novel method to detect C&C communication.

2 Botnet C&C Communication Detection

2.1 Case Study: Rsgh Bot

The rsgh bot is the known IRC-based bot. The rsgh botnet uses the Overnet IRC protocol for controlling the bots. After infection, secondary injection is downloaded from the appointed website. We analyzed the rsgh bot binary and emphasized its C&C communication behaviors.

```
Now talking in #tsgh
Topic is ''
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
XP|00| has joined #tsgh
XP|00| has quit IRC
```

Fig. 1. Rsgh bot joining and quitting the C&C channel

A computer is infected by rsgh bot and then the infected computer will connect the C&C server, which is an IRC server. These bots will join the appointed IRC channel, which is the “rsgh” channel, and wait for the botmaster’s instruction. The control instruction starts with the char “@” and follows with the command, such as “@getclip” which gets clipboard data in the compromised computer. The command type includes getting the system information of the compromised computer, IRC commands, starting server, scanning and speeding, stealing information in the compromised computer, starting ddos attacks, and etc.

Through executing the bot and observing the behavior in the C&C channel, we find that the rsgh bot frequently joins and quits the C&C channel. Figure 1 depicts that the compromised computer uses nick name “XP100ICHN” to join and quit the C&C channel “tsgh”. In order to find the details of the rsgh bot C&C communication, we captured the network packets between the bot and the IRC server in Fig. 2, where the bot’s IP address is 192.168.18.3 and IRC server’s IP address is 192.168.18.2. Figure 2 depicts that the bot joins and quits the C&C channel periodically, and the bot quits and rejoins the C&C channel every one minute thirty-five seconds. Furthermore, we also find an interesting thing that the bot connects the IRC server using different port number every time, as described by Fig.2. The rsgh bot adapts these methods to avoid to be detected, for example Cooke et al. reported that some study detected IRC-based bots if the bots were idle most of time [15], but the rsgh bot is idle for one minute thirty-five seconds the same to the normal human behavior.

Protocol	Local Address	Remote Address	Local Port	Remote Port	Capture Time
TCP	192.168.18.3	192.168.18.2	1423	6667	2011-6-16 7:26:40:420
TCP	192.168.18.3	192.168.18.2	1424	6667	2011-6-16 7:28:14:310
TCP	192.168.18.3	192.168.18.2	1425	6667	2011-6-16 7:29:49:295
TCP	192.168.18.3	192.168.18.2	1426	6667	2011-6-16 7:31:24:295
TCP	192.168.18.3	192.168.18.2	1427	6667	2011-6-16 7:32:59:295
TCP	192.168.18.3	192.168.18.2	1428	6667	2011-6-16 7:34:34:295
TCP	192.168.18.3	192.168.18.2	1429	6667	2011-6-16 7:36:09:295
TCP	192.168.18.3	192.168.18.2	1430	6667	2011-6-16 7:37:44:295
TCP	192.168.18.3	192.168.18.2	1431	6667	2011-6-16 7:39:19:279

Fig. 2. The C&C communication between the rsgh bot and IRC server

2.2 Botnet C&C Communication Detection Model

2.2.1 Botnet C&C Communication Event Assessment

In order to get the botmaster’s commands and finish the tasks immediately, the bots have to often connect the C&C server. Our model accepts as input a packet dump and examines it for the patterns detailed above (as described by section 3.1). In our model, two main parameters control the evaluation of the packets: window and threshold. The window parameter determines the time interval over which we examine for packets (as described by Fig.3) and the threshold parameter determines whether the packet should trigger an alert.

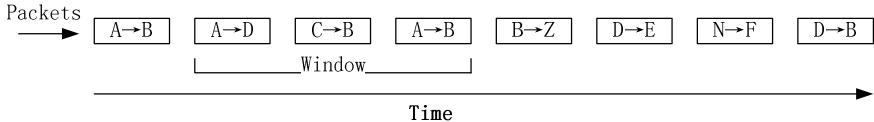


Fig. 3. Packets and checking window

In a window, we can give the score $S(x)$ of a packet x :

$$S(x) = -\log p(x) \quad (1)$$

where $p(x)$ is the probability of the packet for a given port/IP combination x . In order to hidden the bot, the bot has to decrease the communication targeting to the C&C server, and so we can get a higher score than the normal packets. In a window, we can define the total score of the packets between the bot and C&C server as:

$$S(X) = \sum_{t \in T, x \in X} s_t(x) \quad (2)$$

where T is the time window and X is the given port/IP combination. After we count the total score, we can decide whether the bot is communicating with the C&C server:

$$f_{detect}(X) = \begin{cases} 0, & \text{iff } S(X) < \theta \\ 1, & \text{iff } S(X) \geq \theta \end{cases} \quad (3)$$

If the total score is bigger than the given threshold θ , the functions $f_{detect}(X)$ returns 1 and then an alert is triggered. Otherwise, the function $f_{detect}(X)$ returns 0 and then no packets between the bot and C&C server are found.

3.2.2 Botnet C&C Communication Event Correlation

Botnet is an attack network composed of hundreds of millions of compromised computers. The botmaster needs to send its commands to the bots and then bots execute the commands and report the results. Hence, we can observe that bots within the same botnet will demonstrate very strong synchronization in their commands, responses and activities. Based on the behaviors within the same botnet, we can employ several correlation analysis methods to detect the correlation in the botnet C&C communication events. Maybe, the botnet C&C communication event assessment may produce some false positives and we believe that the botnet C&C communication event correlation can efficiently decrease the false positive rate. The correlation within events is evaluated and the evaluation function is:

$$f(e_1, \dots, e_n) = w_1 l_1(e_1, \dots, e_n) + w_2 l_2(e_1, \dots, e_n) + \dots + w_n l_n(e_1, \dots, e_n) \quad (4)$$

Where e_1, \dots, e_n are the events being reported from different netflows within the same time windows, w_1, \dots, w_n are constants, and l_1, \dots, l_n are heuristic evaluation

functions. The heuristic evaluation functions capture knowledge of how events are connected in the botnet C&C communications. The methods by which a heuristic may operate are based on the behaviors of botnets and all heuristic evaluation functions produce results within the continuous range [0, 1], where 0 indicates the heuristic finds no correlation within the events and 1 indicates the strongest possible correlation within the events. Initial botnet heuristics likely include:

Feature source IP address equality heuristics. Is the source IP addressing the same within the events? The source port? If so, 1. Else, 0,

Feature source IP destination equality heuristics. Is the destination IP addressing the same within the events? The destination port? If so, 1. Else, 0,

Feature packet length equality heuristics. Is packet length the same within the events? If so, 1. Else, 0,

Feature proximity heuristics. How close are the times of the events? The closer, the closer to 1 the result would be. Vice versa, the events are too far apart, the result is 0.

The bigger value of the evaluation function $f(e_1, \dots, e_n)$ shows that the events e_1, \dots, e_n are the C&C communications within the same botnet.

3 Conclusions

Botnets pose the most serious threat against cyber-security as they provide a distributed platform for several illegal activities. The C&C communications play an important role in botnets and one effective technique for botnet detection is to identify botnet C&C traffic. In this paper, we analyze the characters of the C&C communication in botnets and then present a novel method to detect the C&C communications. We develop quantitative ways to assess the C&C communications between the bot and the C&C server; furthermore, we also illustrate the correlation methods within the same botnet's C&C communications to decrease the false positive rate. In the future, we will use the method to detect the C&C communications in P2P-based and HTTP-based botnet.

Acknowledgments. This work is supported by special technology development fund for research institutes of the Ministry of Science and Technology of China (2009EG126226, 2010EG126236, and 2011EG126038), China Postdoctoral Science Foundation (20100480074), Supported by the Fundamental Research Funds for the Central Universities (ZYGX2011J069) and NSFC (61103249, 61100150).

References

1. Kang, J., Song, Y.Z., Zhang, J.Y.: Accurate Detection of Peer-to-Peer Botnet using Multi-Stream Fused Scheme. *Journal of Networks* 6(5), 807–814 (2011)
2. Ehrlich, W.K., Karasaridis, A., Liu, D., Hoeflin, D.: Detection of Spam Hosts and Spam Bots Using Network Flow Traffic Modeling. In: Proceedings of the 3rd USENIX Conference on Large-Scale Exploits and Emergent Threats: Botnets, Spyware, Worms, and More (2010)

3. Cui, X., Fang, B.X., Yin, L.H., Liu, X.Y., Zang, T.N.: Andbot: Towards Advanced Mobile Botnets. In: Proceedings of the 4th USENIX Conference on Large-Scale Exploits and Emergent Threats (2011)
4. Feily, M., Shahrestani, A.: A Survey of Botnet and Botnet Detection. In: Third International Conference on Emerging Security Information, Systems and Technologies, pp. 268–273 (2009)
5. Singh, S., Gyanchandani, M.: Analysis of Botnet Behavior Using Queuing Theory. International Journal of Computer Science & Communication 1(2), 239–241 (2010)
6. Honeynet Project and Research Alliance. Know your enemy: Tracking Botnets, <http://www.honeynet.org/papers/bots/>
7. Ramachandran, A., Feamster, N.: Understanding the network-level behavior of spammers. In: Proceedings of ACM SIGCOMM (2006)
8. Choo, K.K.R.: Zombies and Botnets, Trends and issues in crime and criminal justice, vol. 333. Australian Institute of Criminology, Canberra (2007)
9. Freiling, F.C., Holz, T., Wichterski, G.: Botnet Tracking: Exploring a Root-Cause Methodology to Prevent Distributed Denial-of-Service Attacks. In: di Vimercati, S.d.C., Syverson, P.F., Gollmann, D. (eds.) ESORICS 2005. LNCS, vol. 3679, pp. 319–335. Springer, Heidelberg (2005)
10. Barford, P., Yegneswaran, V.: An Inside Look at Botnets. In: Advances in Information Security. Springer (2006)
11. Dagon, D., Zou, C., Lee, W.: Modeling botnet propagation using time zones. In: Proceedings of 13th Network and Distributed System Security Symposium, NDSS 2006 (2006)
12. Gu, G., Porras, P., Yegneswaran, V., Fonk, M., Lee, W.: BotHunter: Detecting malware infection through IDS-driven dialog correlation. In: Proceedings of 16th Usenix Security Symposium, pp. 167–182 (2007)
13. Goebel, J., Holz, T.: Rishi: Identify bot contaminated hosts by irc nickname evaluation. In: Proceedings of 1st Workshop on Hot Topics in Understanding Botnets (2007)
14. Yu, X.C., Dong, X.M., Yu, G., Qin, Y.H., Yue, D.J., Zhao, Y.: Online Botnet Detection Based on Incremental Discrete Fourier Transform. Journal of Networks 5(5), 568–576 (2010)
15. Cooke, E., Jahanian, F., McPherson, D.: The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets. In: Proceedings of the USENIX SRUTI Workshop, pp. 39–44 (2005)

H.264 Decoder Based on Motion Compensation Optimizer

Xiaohong Hao and Jiaqi Li

Electrical Engineering and Information Engineering College,
Lanzhou University of Technology, 69121 Lanzhou, China
haoxiaohong_9@163.com

Abstract. There is a method to reduce the calculation of motion compensation module of H.264 decoder, in order to produce a way to automatically adjust 1/2 and 1/4 pixel precision filter based on image's partial smoothness, in which way it can reduce the drift effect. From the experiment result, the way can effectively reduce the calculation of motion compensation module; in the meanwhile it will not dramatically reduce the video quality and drift effect.

Keywords: Motion compensation, Optimizer, H.264, Drift effect.

1 Introduction

H.264/AVC is a new generation video decoder standard issued by ITU—T and ISO, in order to adapt to the increased video application. There are many new qualities in H.264, for instance, to increase the complexity of calculation aims at high coding efficiency. With improved technology, the calculation complexity of H.264 is higher than previous video coding standards [2-3].

However, with the development of integrated technology and wireless communication, more and more hand-held device enters into household, such as mobile phone, intelligence terminal, and pads. The advantages of these devices are that they are portable, but the downsides are that they have poor CPU calculation capacity, and limited battery storage. H.264's massive calculation has been an obstacle in the spread of these devices. A lot has been done to reduce the calculation complexity of H.264 decoder; document [4] specifically provides an optimizer solution to DM642 DSP. Document [5] provides a solution for ARM Cotex-A8. All the solutions aim as specific hardware platforms for optimization. It makes use of the qualities of the process, such as DMA controller, SIMD order to increase the calculation. Document [6] provides a solution based on random platform, which can effectively reduce the overall calculation. But the downsides are that it will take about four times storage space than before. As for portable devices with limited memory, this solution fails. Besides, mentioned solution will not interfere with video quality after optimization, that is, no damage to the video quality. In reality, the video quality compromise, to some extent, is unnoticeable to naked eyes, but the video playing smoothness and battery duration are more important. Although document [2] has come up with a solution based on decoder's calculation complexity optimizer module,

it needs the decoder to support the decoding complexity controlling function and code stream from average decoder cannot meet the need. Therefore, it is necessary to produce a solution to get low calculation complexity at the cost of certain compromise of video quality; meanwhile as for average code stream from average decoder, there is no need for change of decoder. From document [3], calculations concentrate on several modules, which involve VLC (25% of the total), inverse transform module (15%), motion compensation module (15%), loop filter module (30%). We can optimize these four modules respectively. This thesis focuses on the motion compensation module, in which 70% of the calculation is applied to pixel interpolation filter. There, I propose a pixel interpolation filter calculation to reduce the calculation complexity, in order to effectively decrease the complexity and compromise of image quality. Besides, when GOP is relatively long, there will be no obvious drift effect.

2 The Process of H.264 Motion Compensation Interpolation and Optimization Solution

There are two processes of the pixel interpolation calculation of H.264 motion compensation module, as in picture 1.

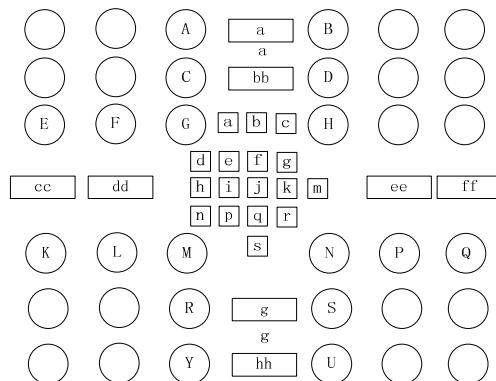


Fig. 1. Interpolation filter

If the pixel position of the image is at 1/2 pixel, dots b, j, h need 6 tap filter to obtain the interpolate pixel. The standardized tap coefficient of H.264 is (1,-5,20,20,-5,1). As for b,

$$b = \text{clip}((E - 5F + 20G + 20H - 5I + J) + 16) \gg 5 \quad (1)$$

In this equation, \gg is displacement calculation, $\text{clip}()$ is calculation restricting variables from 0~255. Calculation for other 1/2 pixel is similar to the previous. As for pixel in position J, first, get the pixel of aa, bb, b, g, gg, hh in a way similar to getting b; then use the same solution to get j's pixel.

If the image's pixel is at 1/4 pixel position, the predicted pixel is the average pixel of nearest two 1/2 pixel position or that of integers' pixel. Take the pixel of a for example

$$a = \text{clip}((G + b + 1) \gg 1) \quad (2)$$

Other predicted values of 1/4 position's pixel can be obtained through similar calculation.

Due to aliasing effect, the interpolation calculation cannot reach 100% precision. The aliasing effect is related to bandwidth of partial image signal. The wider the bandwidth, the stronger the aliasing effect [7-8].

As for average video, most of the images are smooth with narrow bandwidth; therefore the guideline of the calculation is: firstly, as for 1/2 pixel interpolation, low-pass filter with less tap are applied to the smooth area with narrow bandwidth. Because these areas suffer less from aliasing effect, filter with less tap has less effect on the precision of interpolation. Whereas, areas with wide image signal bandwidth, low-pass filter with more tap and counteracting aliasing effect is applied. Applicable tap coefficients are 6 tap, 4 tap, 2 tap (linear interpolation) and 1 tap(nearest interpolation). Besides, as for interpolation calculation of 1/4 pixel, combine the processes of 1/4 pixel interpolation filter and that of 1/2, and get the predicted value of 1/4 pixel by means of asymmetric single filter.

3 Optimization of Interpolation Filter of H.264 Motion Compensation

3.1 The Criteria of Image Signal's Partial Smoothness

The criteria of image smoothness have to meet following conditions: 1) it must be able to precisely reflect the bandwidth and fluctuation of partial image; 2) the calculation has to be minimized, otherwise the optimizer is meaningless.

Take horizontal 4 X 1 pixel module as an example to illustrate how to make sure the smoothness of partial image. As demonstrated in picture 2, \square refers to the pixel of integers, \circ refers to pixel requiring interpolation calculation.

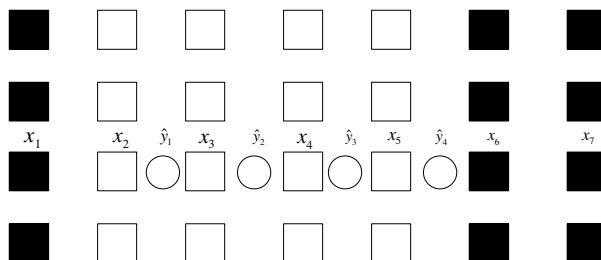


Fig. 2. Interpolation of 4 X 1 as a unit

Assuming the interpolation calculation is going to be conducted on highlighted pixels $x = (x_2 + x_3 + x_4 + x_5)$, firstly, we need to calculate the average of the four points.

$$\bar{x} = (x_2 + x_3 + x_4 + x_5) \gg 2 \quad (3)$$

The partial smoothness of the image is

$$v = \sum_{i=2}^5 |x_{li} - \bar{x}| \quad (4)$$

In the equation: $| |$ refers to absolute value. Through experiment, the distribution of smoothness is illustrated in picture 3.

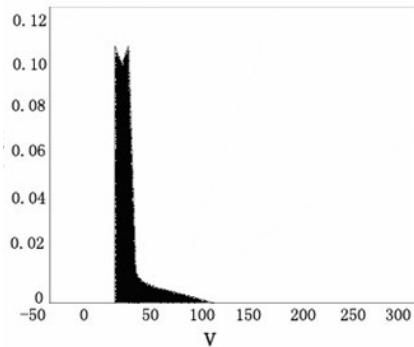


Fig. 3. The distribution of image's smoothness

From picture 3, we can see smoothness of most area focuses on lower area, which goes along with the hypothesis that most area of a image is partially smooth. Therefore, v can be used as a guideline to determine tap numbers. The smaller v is, the smoother the image is and fewer the tap. Otherwise, the bigger v is, corresponding taps are more. When $v=0$, it proves that image signal doesn't change, and the nearest interpolation can be used. Let's set two threshold values V_1 and V_2 ; when v is between 0 and V_1 , tap 2 filter is adopted; when v is between V_1 and V_2 , tap 4 filter is adopted; when v is more than V_2 , tap 6 filter is used. Therefore, the choice of V_1 and V_2 can adjust the relation between decode image quality and calculation volume. The specific choice can be made through experiment.

3.2 The Determination of the Tap Coefficient of the Optimized 1/2 Pixel Interpolation Filter

From the above criteria, we only need to determine the tap coefficient of the optimized tap 4 filter. To reduce the effect of aliasing effect, H.264 adopt Wiener

filter as interpolation filter [9], thus when deducing the optimized tap 4 filter, we can use Wiener filter's design for our own benefit.

As in picture 2, assuming the tap 4 filter is $h = (h_1, h_2, h_3, h_4)$, the interpolation pixel can be

$$\hat{y}_i = \sum_{j=1}^4 h_j x_i + j \quad (5)$$

When $y = (y_1, y_2, y_3, y_4)$ demonstrate the actual pixel, the partial optimized interpolation filter should minimize equation (6).

$$e = \frac{1}{2} \sum_{i=1}^4 \left(y_i - \sum_{j=1}^4 h_j x_i + j \right)^2 \quad (6)$$

Because the tap 4 filter can be only used for image modules 1×4 or 4×1 when v is between V_1 and V_2 ; therefore optimized filter coefficient in overall area should meet the need to minimize equation (7)

$$E = \sum_{V_1 \leq v_m < V_2} e_m \quad (7)$$

In the equation: m refers to number m of training sequence, fitting image modules of 1×4 or 4×1 between V_1 and V_2 . To minimize equation (7), here is

$$\frac{\partial E}{\partial h_k} = 0 \quad (8)$$

Then

$$\sum_{V_1 < v_m < V_2} C_{m,k}^{xy} = \sum_{V_1 < v_m < V_2} \sum_{j=1}^4 h_j C_{m,k+1}^{xx} \quad (k = 1, 2, 3, 4) \quad (9)$$

In the equation: $C_{m,k}^{xy} = \sum_{i=1}^4 x_{m,i} y_{m,i+k}$ and $C_{m,l}^{xy} = \sum_{i=1}^4 x_{m,i} y_{m,i+l}$ refer to signal's cross-correlation function self correlation function respectively. From equation (9), we can get quartic linear equations, whose solution can produce the optimized coefficient of the filter.

When it comes to 1/2 pixels tap 4 filters, there are other conditions. First, the filter should meet the demand of symmetry, that is, $h_1 = h_4, h_2 = h_3$. To ensure the image's brightness in H.264 will not change after filtering, the sum of all the coefficients needs to be 1. These constrains should be brought into the equations before solution.

This thesis chooses some frames from foreman, crew and city video sequences as training sequence. Firstly, de-sampling filter these images and get x, then the actual y can be obtained from the original image. By the above methods, round the tap coefficient and get the tap 4 coefficient of optimized 1/2 interpolation is $(-3, 19, 19, -3)/32$.

3.3 Optimization of 1/4 Pixel Precision Interpolation Filter

In the decoding process of H.264, 1/4 pixel's value is the average value of its two nearest 1/2 pixels' values after calculating their pixels first. These two processes are separate, which leads to unnecessary calculation. As for those points where 1/4 pixel interpolation will be conducted, there two processes can be merged by using one interpolation filter to get the interpolation value of 1/4 pixel directly.

Based on different location of 1/4 pixel, there are differences in specific calculation methods. As illustrated in picture 1, as for 1/4 pixel of point a, we can combine equation (1) and (2) to get the following way:

$$a = \text{clip}(((E - 5F + 52G + 20H - 5I + J) + 32) \gg 6) \quad (10)$$

That is to simplify the original two-step calculation into one filtering process, saving the initial sum and average process. This process can reduce the calculation through reduction of tap numbers. For instance, if the first stage of point a corresponds to 1/2 interpolation filter with tap coefficients (-3,19,19,-3), the calculation of this point is:

$$a = \text{clip}((-3F + 51G + 19H - 3I) + 32) \gg 6 \quad (11)$$

If the corresponding 1/2 interpolation is the nearest one, 1/4 interpolation can use the nearest differential value method. 1/4 pixel in other positions can adopt the same method.

3.4 Reduce Drift Effect

The definition of drift effect is that if the image quality of GOP, as reference frame, is somewhat compromised, then the defect will pass to other frames that use this frame as reference. If future frames use defect frames as reference, the defect of the image will be accumulated, resulting decreasing image quality in GOP. This situation is worse when the GOP is longer. In order to reduce the drift effect, use a little optimization on P frame of GOP, and more on B frame, in which way the drift effect can be reduced to some extent.

4 Experiment's Result and Analysis

This optimization is based on JM8.6 code. To test the probability of the calculation in this thesis, in experiment the code in JM8.6 will not be changed unless it is involved in calculation (code and function, irrelevant to decoding, are erased). The testing condition is based on Simit-Arm software operated on Linux platform. This software is a simulator of 32 ARM process with StrongARM1100 and 32 Mbyte memories. This software can precisely calculate the total instruction numbers while decoding, and ticks' numbers while running on the processor. These two data can be used as standard for the calculation complexity in this thesis. In the trial, the code compiling

device is arm-linux-gcc, and its version is 2.59.3. In compiling, 03 optimization would be adopted unanimously.

Video sequence in the trial uses JM8.6 coding compressor, with 100-frame-length GoP, and the structure of GoP is IPBPBPBP... To test the loss of image quality after optimization, we will use PSNR after decoding 100 frames sequence to demonstrate the image quality. Testing sequences will choose two-cif-distinction video sequences foreman and coastguard, and two D1 distinction video sequence city and crew. All the frame frequency is 30f/s. The above four sequences will be compiled according to QP=28 and QP=32. Considering calculation and image quality loss and drift effect, we will apply V1=50,V2=100 to B frame and V1=20,V2=40 to P frame. In the trial, we only provide calculation aiming to increase optimization of MC module and get its calculation. The complexity of non-optimized calculation is to minus MC module's calculation complexity. The result is shown in chart 1.

Table 1. Comparison between before and after the optimization

sequence	QP	Code rate/(kbit·s ⁻¹)	Reduction of total instruction number/%	Tick reduction /%	Tick reduction of MC module/%	PSRN reduction /dB
Foreman (cif)	28	310.98	7.86	5.11	33.81	0.40
	32	233.54	8.40	5.65	43.16	0.26
coastguard (cif)	28	1015.26	3.96	2.02	14.67	0.27
	32	627.78	4.43	2.74	20.86	0.15
city(D1)	28	1492.22	5.96	3.07	23.11	0.40
	32	884.94	7.66	4.45	28.83	0.27
crew(D1)	28	1968.82	6.13	4.04	29.22	0.37
	32	1357.49	6.92	4.41	28.94	0.25

From the result, we can see the calculation proposed in the thesis improves those with code rate reaching 4% of the total, and the reduction of average PSNR of the total GoP can be controlled under 0.4 dB (1%). Because long 100-frame GOP structure is adopted in the trial, the calculation in the thesis doesn't bring in obvious drift effect as indicated in the trial. However, this calculation is more suitable for video with low-frequency such as foreman and crew, while not so preferable for video with high-frequency such as city and coastguard. All in all, the low-frequency video performs better than that with high-frequency, because low-frequency one contains less high frequency. Besides, the proportion of the reduction of total instruction numbers is normally higher than total tick numbers; it is because the calculation in this thesis mainly aims to reduce the calculation-like instruction, such as sum, multiple and displacement instruction, which takes less tick numbers. In contrast, instructions, such as memory visit while missing cache, take much longer tick numbers. Thus the optimization of instruction numbers is higher than tick numbers in the actual operation.

5 Conclusion

This thesis comes up with a solution to reduce motion compensation module in H.264 decoder, to realize the interpolation filter's calculation of 1/2 pixel and 1/4 pixel precision which can be automatically adjusted by image's partial smoothness and to produce a method to reduce the drift effect. From the experiment's results, the calculation in this thesis can effectively reduce the total calculation complexity of decoding process, without bringing in obvious image quality loss and serious drift effect. There is still more to dig into concerning this topic, such as a way to improve the calculation of high-frequency video and video with more high-frequency part, and a way to optimize the instructions of memory visit taking up more time.

References

1. Wiegand, T., Sullivan, G.J., Bjontegaard, G., et al.: Overview of the H.264/AVC video coding standard. *IEEE Trans. Circuits and Systems for Video Technology* 13(7), 560–576 (2007)
2. Lee, S.W., Kuo, C.C.: Motion compensation complexity model for decoder-friendly H.264 system design. In: Proc. 9th Workshop on Multimedia Signal Processing, pp. 119–122. IEEE Press (2007)
3. Wei, X., Wei, J., Zhang, P.: Embedded platform H.264 software decoder s optimization. *Television Technology* 34(5) (2010)
4. Horowitz, M., Joch, A., Kossentini, F., et al.: H 264 AVC baseline profile decoder complexity analysis (EB/OL) (August 20, 2010),
http://vc.cs.nthu.edu.tw/homelpaper/codfiles/shihyu/200407141306/H.264_AVC_Baseline_Profile_Decoder_complexity_analysis.ppt
5. Pujara, C., Modi, A., Sandeep, G., et al.: H 264 video decoder optimization on ARM Cortex-A8 with NEON (EB/OL) (August 20, 2010)
6. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5409460
7. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. *IEEE Trans. Image Processing* 15(2), 430–444 (2006)
8. Zhang, T., Zhang, Y., Pin, X., et al.: Detection of ILSB steganography based on image smoothness,
http://ieeexplor.ieee.org/xpl/freeabs_a11.jsp?arnumhnr=4036865
9. Wedi, T., Musmann, H.G.: Motion-and aliasing-compensated prediction for hybrid video coding. *IEEE Trans. Circuits and Systems for Video Technology* 13(7), 577–586 (2003)

Influence of Scattering on Laser Communication

Junjun Wu and Mingchuan Meng

Qing Gong College of HeBei United University, Tang Shan, China
IJunjun-yd@163.com

Abstract. Atmospheric laser communications use a laser beam as a carrier of information transmission, atmospheric laser communication technology use atmosphere as the transmission medium to transmit voice, data, images and other information in the space directly. As a flexible and convenient new communications technology, laser communications has become a hot topic in today's world communication technology and has very broad prospects for development, but the quality of the laser communication is influenced by atmospheric channel, atmospheric attenuation refers to the weakening of the laser signal energy due to atmospheric absorption and scattering. Paper analyse the influence of scattering on laser communication, getting the relation between laser signal attenuation and transmitting distance, work wavelength and visibility. The laser signal attenuation is simulated from different wavelength, different transmitting distance and different visibility. It turns out that the signal degradation caused by atmosphere scattering is stronger as the work wavelength turn longer, laser signal attenuation decreases with the increase of visibility, and the signal degradation improve as increasing transmitting distance, so choosing long work wavelength, short transmitting distance and good visibility can improve the quality of laser communication.

Keywords: Scattering, Attenuation, Numerical Simulation, Laser Communication.

1 Introduction

As a flexible and convenient new communications technology, Laser communication technology has been the hot topic of world communication technology development, the developing trend and potential of laser communication technology attach great importance to many countrys. Atmospheric laser communications use a laser beam as a carrier of information transmission, atmospheric laser communication technology use atmosphere as the transmission medium to transmit voice, data, images and other information in the space directly. However, laser transmission in the atmosphere will be its attenuation of the atmospheric channel, including the scattering of atmospheric molecules and aerosol particles, absorption, and large air flow, so that the laser communication signal attenuation, the paper carry out research from an angle of atmospheric molecules scattering and the air channel of the laser signal to improve the quality of the laser communication [1-5].

2 The Scattering of Atmospheric Aerosol Particles

Earth's atmosphere is the composition of a variety of gases and suspended solid particles or liquid particles (known as atmospheric aerosols). The gas composition of the lower atmosphere can be divided into three categories: one is essentially the same components, including nitrogen, oxygen, hydrogen, As well as trace amounts of inert gases as neon, helium, fluorine, atmosphere, etc. their average life expectancy is greater than 1000a ,they hold a fixed proportion in atmospheric composition mediator ; second category is a variable composition, the ratio changes as time, location changes, water vapor increase the most, carbon dioxide and ozone is a small proportion of atmosphere, but they have a major impact on the transmission of light waves; The third category is gas composition that changing rapidly and their average life expectancy is shorter than one hour, such as carbon, sulfur, and nitrogen compounds. Usually the pure atmosphere in addition to water vapor is known as the dry clean air, referred to dry air. Its main ingredient is nitrogen, oxygen, hydrogen, carbon dioxide, and so on.

The diameter distribution of the various particles in the atmosphere is very wide, from 4 microns to tens of microns. The greatest impact of the laser signal transmission is the particle diameter from 0.1 microns to 10 microns, which can be used single-atom scattering theory of Rayleigh scattering and of mie scattering to approximate analysis the atmospheric scattering. This is because when the spacing of gas molecules is greater than the molecular diameter of more than ten times, the atmospheric aerosol particles or aerosols spacing is much larger than the particle diameter to meet the conditions of the single-particle scattering.

2.1 Rayleigh Scattering Coefficient

When the wavelength of laser signal is much longer than the size of scattering particle, Rayleigh scattering happen, scattering coefficient is

$$\alpha_n = \frac{32\pi^3(n-1)^2}{3N\lambda^4} \quad (1)$$

In the formula, N is the amount of particle in unit volume. K is wavelength; n is refractive index of medium, in the clearness air, rayleigh dispersion coefficient

$$\alpha_n(\lambda) = 1.09 \times 10^{-3} \lambda^{-4.05} km^{-1} \quad (2)$$

The main features of Rayleigh scattering are:

- (1) The power of scattered light is inversely proportional to the fourth wavelength.
- (2) A relatively simple relationship between the scattering intensity of scattered light and observation direction.
- (3) Forward energy of scattering is equal to the backward scattering energy.
- (4) The scattered light of 90° direction of is almost fully polarized. In general, the particle of the radius $r \leq 0.03 \text{ Lm}$, the wavelength is near 1 Lm, the error of Rayleigh scattering is smaller than one percent [6-11].

2.2 Mie Scattering Coefficients

When the size of dispersion particle is close to the wavelength, the mie dispersion produce which is the same with such as rain, mist, haze and so on. The main features of the mie scattering:

(1) scattering light intensity becomes very complex with the angle distribution. The larger the relative wavelength scale of particle, the more complex the distribution.

(2) the ratio of forward scattering and after scattering increases when the scale of the particle increase.

(3) With the further increase in particle size, the angle of scattered light concentrated is more and more narrow [12].

Mie dispersion coefficient can be determined by the follow formula:

$$\alpha_n = \pi N(r) Q_s(X_r, m) r^2$$

In the formula, N is the amount of particle in unit volume; r is radius of particle; $X_r = \frac{2\pi r}{\lambda}$, refractive index $m = n - iK_a$, Q_s is scattering efficiency which is defined the ratio of dispersion energy to incidence energy, it is relative function of particle, Ka and n is imaginary part and real part of refractive index respectively. When the radius of particle is continuum between r_1 and r_2 , mie dispersion coefficient can be getting from the follow integral:

$$\alpha_m = \pi \int_{r_1}^{r_2} n(r) Q_s(X_r, m) r^2 dr$$

in the formula, n(r) is the number of particle radius between r and r+dr.

3 The Influence of Atmosphere Molecule Scattering on Laser Transmitting Characteristic

To wireless communications, the scattering of atmosphere lead to the energy of the laser signal generated attenuation, strength occur up and down, impulse broadening. The reason for laser impulse being broaden mainly is multipath effect of laser transmission caused by multiple scattering of atmosphere, atmosphere scattering make laser transmitting power attenuation, and signal noise ratio receiver detected decrease, laser power density receiver detected decrease, and form so called atmospheric extinction which is atmosphere attenuation. The essence of atmosphere dispersion is changing the transmitting direction of light wave. The factual atmosphere contain dispersion particle of different size, each particle independently bring dispersion to laser beam, the summation of many special dispersion form multi dispersion. When laser transmitting signal pass atmosphere random channel, interaction between all

atmosphere particle and all laser radialization field that through radialization zone happen, but most of basic effect is atmosphere attenuation.

The light power transmitting in atmosphere can be expressed by the flow formula [12]

$$P(L) = P(0) \exp(-\sigma L) \quad (4)$$

In the formula, $P(0)$ is emission light power, $P(L)$ is light power after laser transmitting L (Km), σ is attenuation coefficient of atmosphere channel.

$$\sigma = \sigma_m + \sigma_\alpha + \beta_m + \beta_\alpha$$

σ_m is molecule absorbing coefficient, σ_α is aerosol absorbing coefficient, β_m is coefficient of molecule rayleigh scattering or molecule mie scattering.

It is obvious that the atmosphere attenuation mainly involves dispersion and absorbing, in a general way, atmosphere molecule absorbing is secondary in atmosphere attenuation.

Atmosphere has different transmission ratio to different beam of light and has many communicate window which transmission ratio is almost 100%. When we design laser communication system, we can choose the work wavelength which is in the scope of channel window, ignoring the power attenuation, $\sigma_m = 0$. rayleigh scattering coefficient is in inverse proportion to quartic wavelength, and rayleigh scattering coefficient is very small in infrared ray, so β_m also can be ignored, the rock-bottom atmosphere attenuation mainly involve mie scattering.

The light power after laser transmitting distance L :

$$P(L, r) = P_0(L) \exp\left[-\frac{2r^2}{\omega^2(L)}\right]$$

The connection between $P_0(L)$ and P_s is:

$$P_s = \int_0^{2\pi} \int_o^{+\infty} P_0(L) \exp\left[-\frac{2r^2}{\omega^2(L)}\right] r dr d\varphi = \frac{\pi \omega^2(L)}{2} P_0(L)$$

If receiving aperture is D , and is perpendicular to the direction of propagation of the laser beam, then the received optical power is:

$$P_r(L) = \int_0^{2\pi} \int_o^{\frac{D}{2}} P(L, r) r dr d\varphi = \frac{\pi \omega^2(L)}{2} p_o(L) \{1 - \exp\left[-\frac{D^2}{2\omega^2(L)}\right]\}$$

We can derive the loss of laser beam power after transmission L :

$$\alpha = 10 \log \frac{1}{1 - \exp\left[\frac{D^2}{2\omega^2(L)}\right]} \quad (5)$$

4 Numerical Simulation

From the above theoretical analysis, the attenuation of laser signal energy influence by the atmospheric channel, the scattering of atmospheric molecules accounted for large part of reason. The degree of signal attenuation has relations with weather visibility, transmission distance, wavelength, the receiving aperture and other factors. Based on the above theoretical model, Making numerical simulation using matlab software in order to get the relationship between the degree of signal attenuation and each factor, we have three analogs of attenuation simulation, the different visibility, attenuation simulation of different transmission distance, attenuation simulation of different operating wavelengths.Fig.1 is the relation among signal degradation and visibility when transmitting distance is two kilometers and work wavelength is 850nm.we can see that the laser signal attenuation is less as the accretion of visibility. Fig.2 is the relation among signal degradation and work wavelength when transmitting distance is certain in clearness weather, we can see that the long wavelength is profitable to laser communication.Fig.3 is the relation among signal degradation and transmit distance when the waist radius is same and the different work wavelength is different, it show that the longer transmitting distance is, the stronger signal degradation.

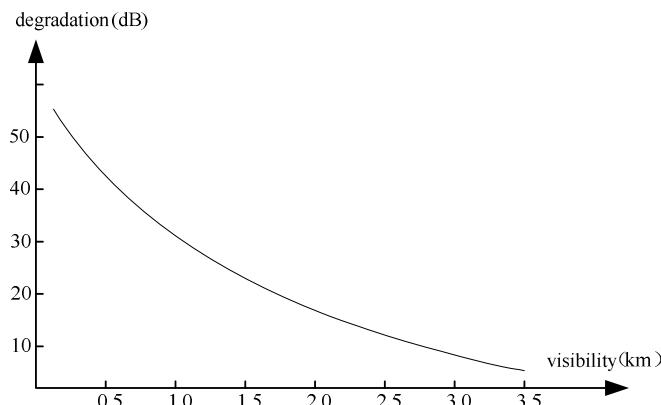


Fig. 1. The relation between signal degradation and visibility

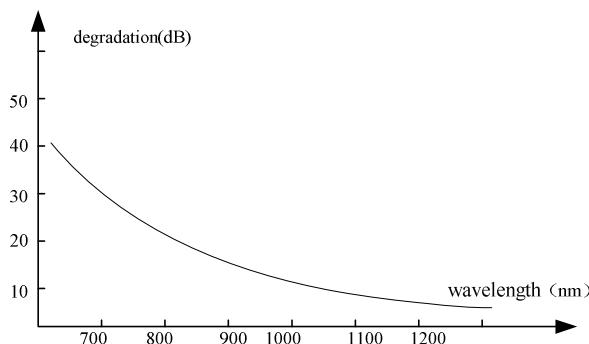


Fig. 2. The relation of signal degradation and work wavelength

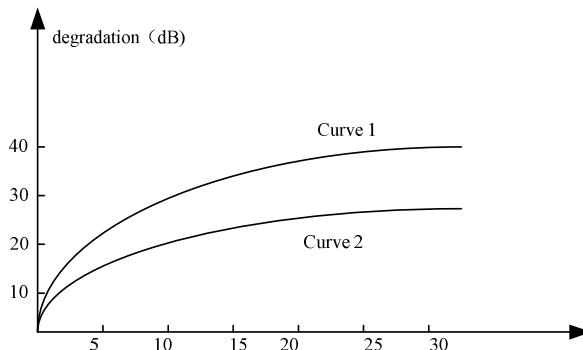


Fig. 3. The relation of signal degradation and transmitting distance Curve 1 $\lambda = 850\text{nm}$ $\omega_0 = 4\text{cm}$ Curve 2 $\lambda = 1550\text{nm}$ $\omega_0 = 4\text{cm}$

5 Conclusion

The quality of the laser communication is influenced by atmospheric channel, there are numerical modeling to laser signal from different transmitting distance, different work wavelength and different visibility, It turns out that the signal degradation caused by atmosphere dispersion is stronger as the work wavelength turn longer, laser signal attenuation decreases with the increase of visibility, and the signal degradation improve as increasing transmitting distance, so choosing long work wavelength, short transmitting distance and good visibility can improve the quality of laser communication.

References

1. Yuan, Z.H., Zhang, W.T.: Analysis of the laser beam through random atmosphere channel. Chinese Journal of Quantum Electronics 23(5), 696 (2006)
2. Li, H.: The principle and system of optical communication, pp. 225–226. High Education Publishing House (2009)
3. Ren, G.: Key technology and its application of atmospheric laser communication. Science and Technology Information, Hei, L.(2010)
4. Corrigan, P., Martini, R., et al.: Quantum cascade lasers and the Kruse model in free space optical communication. Optics Express 17(6), 4355–4359 (2009)
5. Wang, L.-L., Ke, X.-Z., Xi, X.-L.: Analyze and Calculate for Laser Transmit Through Rain. Chinese Journal of Light Scattering (2) (2005)
6. Wang, H.-X.: Technology of laser communication in atmosphere. Infrared and Laser Engineering 30(2), 125 (2001)
7. Song, X.-P., Jia, X.-D.: Experiment research on laser backscattering characteristic of fog. Infrared and Laser Engineering 39(2), 261–265 (2010)
8. Sun, J.-Q.: Atmosphere Detecting by Lidar. Science Press, Beijing (1986)
9. Liu, J.-B., Wu, J.: Light scattering model and angular spectrum of scattering intensity in fogs

10. He, X., Yin, Z.: Coding theory in wireless laser communication system. Science Press, BeiJing (2009)
11. Zhang, W., Zhu, B.: The study for laser beam transmission character in atmosphere random channel by Using CCD. Applied Optics 25(5), 34–35 (2004)
12. Geng, W.: The influence factors and performance analysis of Wireless Optical Communication Systems. Master's thesis of Xi'an University of Electronic Science and Technology (2010)

Research on University Library Services in Science and Technology Novelty Search

Lei Mei-Shen¹ and Zhang Tian²

¹ Library of Hebei United University, Tangshan, Hebei, China

² School of Management of Hebei United University, Tangshan, Hebei, China

{meishenliu, comztcyy}@163.com

Abstract. This paper expounds several main definitions of science and technology novelty search, analyzed the importance and functions of science and technology novelty search in university libraries, and discussed briefly some basic methods. Improvement of continue education for novelty searching personnel in university libraries.

Keywords: University Library, sci-tech novelty retrieval, information retrieval.

1 Three Main Definitions of Science and Technology Novelty

China's science and technology novelty (hereinafter referred to as the "novelty") which got started in the middle of nineteen eighties and developed gradually thereafter is a profound scientific and technical information advisory service. Almost at the same time, novelty search in the University Libraries nationwide was also developed synchronously. In order to regulate the activities of such novelty search organizations, as well as to guarantee the novelty fairness, accuracy and independence, the State Ministry of science and technology in December 2000 issued "science and technology novelty search organization management approach" and "science and technology novelty standard". So what exactly is "novelty"? Although it is still divergent in the public opinions, actually all opinions are almost the same. Integrating its connotation and denotation, it is commonly accepted that novelty includes the following three aspects. First, science and technology novelty search is an information consulting service provided by an agency which possesses certain kind of information resource. Qualified personal acquire literature review results, using comprehensive analysis and comparative method, through computer searches and manual retrieval approaches. This kind of information will be used for the evaluation of novelty of achievements in scientific researches and scientific research proposal¹. Second, science and technology novelty search verifies the novelty of science and technology projects based on information collected and summarized from published literatures. This is mainly performed by information service agencies. Third: Science and technology novelty is the research on information obtained through literature search and information analysis. The fundamental portion of novelty search is dependent on literatures. A research report usually consists of existing evidences,

analyses, comparisons and conclusions based on document retrieval, comprehensive information analysis, as well as review results of information provided in proposed project.

2 Significance of Launching Science and Technology Novelty in University Libraries [2]

2.1 To Ensure the Seriousness of the Research Reviewed, Impartiality

Novelty has changed the mode of scientific research management in colleges which simply relied on the assessment made by professors and experts. It provides professors and experts with comprehensive and accurate literature evidences, which ensures the seriousness, fairness, accuracy and authority for scientific research evaluation. Thus, novelty plays a decisive role in college scientific research managements in China today.

2.2 Provide the Basis for the Identification of Teaching and Research Staff to Apply Scientific Research Projects and Scientific Research

Novelty can provide an objective foundation for research project applications, as well as appraisal, assessment, acceptance, transformation, award declaration and patent application of research achievements to teaching and research personnel in universities. At the same time, novelty can also provide instant, reliable, and abundant information to teaching and scientific research personnel for science research, development, and key project overcoming. All these functions fully confirm that novelty plays an important role in scientific research.

2.3 Accelerate Research Speed and Improve the Level of Scientific Research

Since analysis and comparison are performed throughout the whole process, novelty searching can leach out redundant and repetitive scientific research projects, in turn, to remind and help college teaching and scientific research personnel to promptly adjust their scientific research direction. At the same time, it can also choose the most effective and useful information from the massive information database for them to accelerate their scientific research and elevate their scientific research level.

2.4 To Fill the Gaps in the Industry, Improve the Social and Economic Benefits

Through the novelty search, blank, strong or weak aspects can also be found clearly in various scientific fields, which will help turn outstanding achievements in scientific research into productive forces quickly, effectively and fill industry blank promptly and thus create both good social benefits and economic benefits.

3 Effects of Scientific and Technological Novelty Search in University Libraries [3, 4]

Novelty Search can provide experts and professors with systematic, objective, and accurate retrieval of scientific documents and conclusions of informative evaluation for the assessment of science and technology projects. Thus, it can help them improve the accuracy of their evaluation effectively and provide policy-making basis for the scientific research management in colleges.

3.1 To Provide an Objective Basis for Scientific Research Projects

Novelty is required for hypotheses and objectives of a research project as well as its technical strategies, contents, goals, and levels and other aspects. Therefore, before the official establishment of an object, one must collect and review related information comprehensively and accurately globe wide. Through novelty search, the status of such project can be verified as if it has already been investigated or if it is still under investigation and if so, what the depth, and breadth are, what kind of problems have been solved and what the remaining problems are domestically and abroad. These kinds of information will provide objective evidences for the assessment of the novelty of the chosen topic and prevent efficiently the repeat of research and development and waste of human, material and financial resources caused by teaching and scientific research personnel and guarantee limited research money can be used on "the edge of knife".

3.2 To Provide Objective Basis of Literature for the Appraisal of Scientific and Technical Achievements, Assessments, Acceptance, Transformation, and Declaration of Award

Novelty can maximally ensure the scientific validity and reliability of scientific research achievement appraisal, assessment, acceptance, transformation, and reward. Mean well, reliable, efficient, and highly qualified novelty reports provided by university libraries can fundamentally avoid the occurrence of imprecise conclusions due to solely relying on professional knowledge and experiences of professors' and experts' and some injustices. They can also protect and arouse the enthusiasm of teaching and scientific research personnel in universities.

3.3 To Provide Effective, Reliable, Highly Qualified and Abundant Information to Teaching and Research Staff In Universities to Carry Out Scientific Research Activities

With the rapid development of science and technology, massive creation of information from a variety of different carriers has become a common phenomenon. This has brought considerable difficulties for teaching and research staff in universities to obtain useful and reliable information. Novelty search performed by

university professionals can not only save teaching and scientific research personnel a large amount of time on literature consulting, but also provide comprehensive primary and secondary literature services because of possessions by universities of both abundant information resources and a perfect computer retrieval system.

4 Fundamental Novelty Searching Methods Launched by University Libraries [5-8]

Novelty retrieval personnel in university libraries should operate according to the "science and technology novelty standard" to determine the novelty of the entire content of the project based on the requirement proposed by novelty searching client in "project of science and technology novelty search proxy" and make a scientific, objective conclusion.

4.1 To Handle the Contradiction between Objective Documents and Subjective Judgments Properly

Novelty search is a kind of service based on literature resources. Novelty retrieval of relevant literature is objective, while the understanding, analysis, comparison, and integration of literature resources are subjective. However, novelty should be an objective logic judgment because of its characteristic function of authentication. Therefore followings are required: (1) maintain an objective and fire position; (2) be as faithfully as possible to the original expression pattern of obtained literature sources and to avoid arbitrary addition or deletion to the original literature expression pattern; (3) read closely related literatures carefully; (4) refer to as many review articles as possible; (5) express technically as purely as possible; try to use perfect technical terms rigorously, and use different expression patterns for technical terms; (6) Technical terms is as rigorous perfect as possible, do not miss the variety of technical terms of expression; (7) the patent literature should be based on patent claims, at the same time full patent text and abstract should be included [9].

4.2 Must Grasp Correct Novelty-Searching Points

When novelty search project is accepted, University Library novelty retrieval personnel should determine whether the client has clearly shown the novelty search points, listed Novelty-searching points have single themes and there are accurate, and the client fully understands that the project is different from others at the level of technology or application. In addition, novelty retrieval personnel need to pay attention to see if novelty points can be effectively broken down, because it will closely affect the novelty judgment. Sometimes, some clients hasten to give publicity to their research content due to various reasons. In this kind of circumstances, retrieval personnel should explain related rules and the principle of confidentiality to the clients so that they can understand correctly and grasp the point of novelty search [10].

4.3 Utilization of Accurate Retrieval Key Words

The selection of correct retrieval key words is crucial for the screen of all the documents that related to the novelty-searching project from massive information resources. Retrieval words should cover all the technical terms which are publicly known and used in related industries. At the same time, possible synonyms, abbreviations of generic terms, symbols and other forms should be taken into consideration so that omission of retrieval can be avoided. In the English database retrieval, particular attention should be payed to all the transforming patterns of English words such as the part of speech, tense, voice, and number types. Full use of truncation retrieval which can cover all kinds of suffixes is an effective retrieval method. Recall rate should also be guaranteed while precision ratio is in consideration[11].

4.4 To Determine a Reasonable Search Term

Improper retrieval term can sometimes lead to a retrieval result of zero. According to "Science and technology novelty standard ", retrieval term is chosen generally from the date when the agreement was signed and date back for more than 10 years. But in practice, it should be determined mainly by the progressing status of the specialty in which the retrieval project belongs. The prospective period should be shortened for some new or high-tech projects. It has been showed that usually retrieval of last 5 years is likely sufficient. However, the prospective period should be prolonged if the retrieval search is for some well established technological products or manufacturing techniques [12].

4.5 Principles for the Selection of Related Literature

It is an essential step to search for literature resources that are related to the contents of novelty projects. Literatures that are most closely related, comparatively related, and generally related to novelty projects and novelty pinpoints are collected mainly through logical analogy. These documents will be organized and utilized according to the novelty search conclusion.

4.6 The Use of Databases for Novel Search

Databases should be chosen comprehensively and all possibly related information resources should be considered as far as possible. In principle, following points are usually included: (1) Choosing main professional and important databases that are closely related to novelty projects; (2) selecting large, well-known integrated databases; choosing the choice of professional core journal databases; (3) Select the professional core journals database; (4) selecting integrated, well-known search websites, such as: Baidu search engine (www . Baidu. Com) Chinese search Alliance (www . ZhongSou.com) Yahoo (www . Yahoo.com), Google (www . Google. Com). In addition, patent databases can also be used as one of the key retrieval

database, because illustrations for a product or equipment in the patent instruction are an excellent supplement to technical details.

4.7 Composing Novelty Search Conclusion

The ultimate objective for novelty search lies in the novelty retrieval conclusion of the report, and novelty search conclusion is just an evaluation for the novelty of a proposed project [13].

5 Improvement of Continue Education for Novelty Searching Personnel in University Libraries [14]

Novelty retrieval personnel in university libraries should participate in professional trainings and seminars related to novelty through various sources to broaden their knowledge. They should also try to learn new retrieval techniques, and participate in technical exchanges often with colleagues. They should also be familiar with retrieval methods and skills for a variety of search engines especially professional ones, as well as skills for information screening, assessment, and filtration. At the same time they should learn how to summarize from their experiences in time and collect feedbacks and evaluations for novelty projects so that they can improve their own quality continuously in all aspects. The advantage for novel search in university libraries is that their clients are teachers from same universities with research tasks and the contents of their novelty search projects are usually related to various specialties in same universities. At the same time, with this advantage, a consulting team consisted of teachers from different majors can be established and these experts can help novelty retrieval personnel with proper screening, extraction and analysis comparison of selected documents so that They can compensate for novel search professional who lack expertise in order to make an objective and accurate novelty search conclusion. Therefore, to keep sustainable development in service and management in the further, university library novelty search centers have to maintain and launch brand services, and expand profound information retrieval service, which will be the future development direction for sci-tech novelty retrieval station.

In conclusion, the composition of novelty search report must strictly comply with the " science and technology novelty standard" requirement and reflect only the novelty assessment conclusion. Nowadays, with the rapid development of science and technology, the importance and bond effects of science and technology novelty search in the research, development and production and other activities are becoming increasingly apparent in Chinese universities. University libraries, with their abundant information resources, excellent human resources, are the main institute for science and technology novelty search in universities. Sci-tech novelty retrieval rooted in the rich soil fertilized with teaching and research activities in university libraries will certainly produce the sweetest fruits.

References

1. Luo, L., Ma, Y.: Reflection of science and technology novelty search work in universities. *Library work in Colleges and Universities* 27(5), 53–57 (2007)
2. Wang, L.: Science and technology novelty search services correctly guide the scientific research innovation. *Modern Information* 30(7), 80–82 (2010)
3. Xing, Y.L.: Significance and role of Sci-tech Novelty Retrieval in University Libraries. *Library and Information Sciences in Agriculture* 21(12), 150–152 (2009)
4. Jun, E.L.: University library science and technology of sci-tech novelty search service survey and analysis. *Journal of Information* 31(1), 181–184 (2012)
5. Wang, C.H.: Analysis of subject librarian assisted university novelty search service mode. *Journal of Hubei Normal University (Natural Science Edition)* 31(2), 66–69 (2011)
6. Bu, H.J., Ren, G.X.: Role Of Sci-tech Novelty Retrieval in university scientific research. *Library and Information Work* (2), 121–123 (2008)
7. Hou, L.L.: Improving the quality of sci-tech novelty retrieval method research. *Information Research* (9), 90–91 (2010)
8. Wu, M.Q., Fu, L.Y.: A Summary of research on scientific novelty search. *New Century Library* (4), 40–42 (2010)
9. Chen, Z.H., Zhang, W.D.: Construction of performance evaluation model of sci-tech novelty retrieval and Its application. *Library and Information Service* 54(7), 101–104 (2010)
10. Liu, H.: Factors analysis on Influencing sci-tech novelty retrieval quality and how to determine appropriate novelty retrieval points. *Journal of Modern Information* 30(10), 144–146 (2010)
11. Zhang, H.Q., Guo, Q.L., Gao, Y.J.: Probe into the Important factors Influencing the quality of the sci-tech novelty retrieval. *Journal of Modern Information* 30(10), 147–149 (2010)
12. Huang, Y.H.: Practice and exploration of sci-tech novelty retrieval. *Library Development* (10), 77–79 (2010)
13. Xing, Y.L.: Analysis on the measures and factors that affect the quality of retrieval conclusion. *Library Work and Study* (5), 96–98 (2010)
14. Wang, C., Zhou, J., Wu, D.-M.: Science and technology novelty search work in university science and technology innovation activities. *Library and Information Work* 53(53), 21–24 (2009)

Analysis of Science and Technology Activity Based on DEA Model

Wenhai Gao and Lina Yuan

School of Economic and Management, Hebei University of Science and Technology,
Yuhua East Road, 050018 Shijiazhuang, China
gao72hai@163.com

Abstract. The science and technology activity plays a key role in the economic development. The method of data envelopment analysis has advantage over integrated assessment methods on assessing multi-input and multi-output complex system. The science and technology activity is an example of multi-input and multi-output system with different input-output dimension. To describe a new data envelopment analysis (DEA) model evaluated the efficiency of the science and technology activity. To modify the standard DEA model by taking into account the series relationship of the two stages within the whole process. Under their framework, the efficiency of the whole process can be decomposed into the product of the efficiencies of the two sub-processes.

Keywords: Data Envelopment Analysis (DEA), Evaluation Model, Science and Technology Activity, Efficiency.

1 Introduction

DEA is used to establish a best practice group among a set of observed units and to identify the units that are inefficient when compared with the best practice group. DEA also indicates the magnitude of the inefficiencies and improvements possible for the inefficient units. Consider n DMUs to be evaluated, DMU_j ($j = 1, 2, \dots, n$), that consume the amounts $X_j = \{x_{ij}\}$ of m different inputs ($i=1,2,\dots,m$) and produces the amounts $Y_j = \{y_{rj}\}$ of r outputs ($r=1,\dots,s$). Each DEA model seeks to determine which of the n DMUs define an envelopment surface that represents best practice, referred to as the empirical production function or the efficient frontier. Units that lie on the surface are deemed efficient in DEA, while those units that do not are termed inefficient. DEA provides a comprehensive analysis of relative efficiencies for multiple input–multiple output situations by evaluating each DMU and measuring its performance relative to an envelopment surface composed of other DMUs. Those DMUs forming the efficient reference set are known as the peer group for the inefficient units. As the inefficient units are projected onto the envelopment surface, the efficient units closest to the projection and whose linear combination comprises this virtual unit form the peer group for that particular DMU. The targets defined by the efficient projections give an indication of how this DMU can improve to be efficient.

The science and technology resources play an important role in the development of society and economy. Due to the lack of resources, we must take full advantage of the limited resources. Quantitative analysis of some data can only reflect the results and can't track out the reasons. This is the main shortage for evaluation of science and technology resources of industrial enterprises.

In this paper, we focus on the correlation analysis, principal component analysis and data envelopment analysis (DEA) method to propose a two-stage envelopment analysis (DEA) model, which can evaluate the efficiency of science and technology resources and reflect the independence of input element and the concentration of output element.

2 Two-Stage DEA Model Analysis

As introduced in Section 1, consider a two-stage process shown in Figure 1. Suppose that we have n DMUs. We assume that each DMU_j ($j = 1, 2, \dots, n$) has m inputs to the first stage, x_{ij} ($i = 1, 2, \dots, m$) and D outputs from the first stage, z_{dj} ($d = 1, 2, \dots, D$). These D outputs then become the inputs to the second stage and are called intermediate measures. The outputs from the second stage are y_{rj} ($r = 1, 2, \dots, s$). Based upon the CCR model, the efficiency scores of the two-stage process and the two individual stages can be expressed as (1).

$$\theta_j = \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}}, \quad \theta_j^1 = \frac{\sum_{d=1}^D w_d z_{dj}}{\sum_{i=1}^m v_i x_{ij}} \quad \text{and} \quad \theta_j^2 = \frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{d=1}^D \tilde{w}_d z_{dj}} \quad (1)$$

Where v_i, w_d, \tilde{w}_d , and u_r are unknown non-negative weights. Noted that w_d can be equal to \tilde{w}_d .

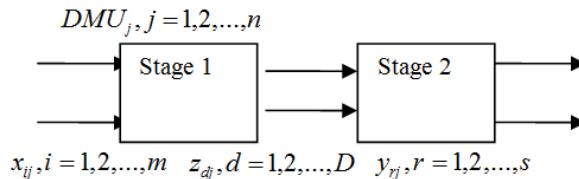


Fig. 1. Two-stage process

We note that both approaches assume that intermediate measures are the only inputs to the second stage. This imposes some limitations of these two approaches in use. It is likely that inputs (not as the outputs from the first stage) exist for the second stage, for example, in a supplier-manufacturer supply chain structure where the manufacturer (as the second stage) can have its own inputs. This can be a further research on how to extend the two approaches to deal with such situations.

3 Establishment of Evaluation Index System

The two-stage DEA evaluation model of the regional science and technology activity efficiency involves a great deal of information. It focuses on establishing a scientific, standardized, systematic, standardization system of index that can comprehensively and objectively reflect the whole process of science and technology activity, sequentially we can have a comprehensive evaluation about the efficiency of science and technology activity.

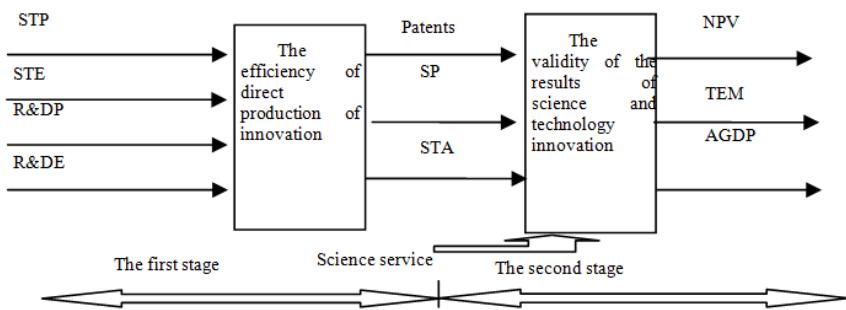


Fig. 2. The two-stage DEA evaluation model of science and technology activity efficiency

It should be noted that the data is sample, because the choice of indexes is based on statistical method. If statistical data is reliable and comes in different years, the choice of indexes should be the same.

4 Application Example

4.1 The Time Series Data Analysis of the Efficiency of Science and Technology Activity

The efficiency of scientific and technological activities is relatively low from 2006 to 2010 in Hebei Province. It can be concluded that the efficiency and investment of scientific and technological activities are high relevant through the analysis of Hebei Province.

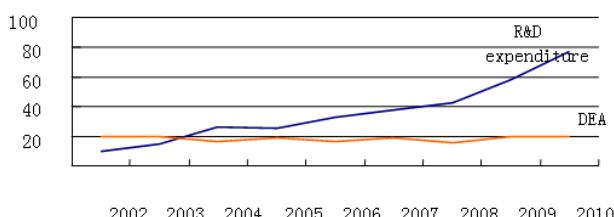


Fig. 3. The relation of R&D expenditure and DEA evaluation

The efficiency of DEA had fluctuated while the R&D expenditure of Hebei Province did by the year of 2003 to 2009. Therefore, it has strong correlation between the STE, R&D and the effectiveness of DEA. The reason is that the investment in science and technology includes scientific and technological personnel, scientific and technological projects, scientific and technological achievements and so on. Investment in science and technology plays a key role to improve the efficiency of scientific and technological activity.

4.2 Economic Impact Analysis of the Efficiency of Scientific and Technological Activities

The regression model is created to test indicators of efficiency of scientific and technological activities. The dependent variable is GDP, and variables conclude the TEM, STP and STA. It is shown as (2).

$$AGDP = \alpha_0 + \alpha_1 TEM + \alpha_2 STP + \alpha_3 STA + \varepsilon_i \quad (2)$$

Where α_0 , α_1 , α_2 , α_3 are coefficient for the model to be estimated, ε is random perturbation.

TEM(technical exchange market);

STP(scientific and technical personnel);

STA(scientific and technological achievements).

Table 1. Model summary a predictors: (constant), TEM, STP, STA

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.829(a)	.688	.652	8318.28390

It can be shown from figure 4 that the relevance is 0.829 between the scientific and technological personnel, the scientific and technological achievements, the technical exchange market and the AGDP, which of determined coefficient is 0.688.

Table 2. ANOVA(b). a Predictors: (Constant), TEM, STP, STA. b Dependent Variable: AGDP

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3965157272.957	3	1321719090.986	19.102	.000(a)
	Residual	1799040024.910	26	69193847.112		
	Total	5764197297.867	29			

The test results of variance F is 55.474, and its corresponding probability is 0.000 <0.05. As a result, at the same time the original assumption of all coefficients on the zero can be refused. Therefore, the variables and the dependent variable is significant linear relationship.

Table 3. Coefficients (a) a Dependent Variable: AGDP

Model	Unstandardized Coefficients			Standar-dized Coeffi-cients		95% Confidence Interval for B		
	B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	
1	(Con-stant)	9361.545	2766.214	3.384	.002	3675.511	15047.579	
	TEM	.006	.001	.561	4.270	.000	.003	.009
	STP	.019	.021	.140	.922	.365	-.024	.062
	STA	8.623	3.348	.338	2.575	.016	1.741	15.506

The regression equation is established in the basis of information.

$$y\wedge=0.006 \text{ TEM} +0.019 \text{ STP} +8.623 \text{ STA} +9361.545 \quad (3)$$

It can be shown that the above assumptions are correct. The impact is significant for the scientific and technological personnel, scientific and technological achievements and technical exchange market to the economic development.

5 Conclusions

This paper introduces a new two-stage DEA model to evaluate systematic efficiency for the science and technology activity. Specifically, this new model allows analysis of the input and output for science and technology activity. Throughout the study, special emphasis was placed on how to present the DEA results to management so as to provide more guidance to them on what to manage and how to accomplish the changes. The results show that the science and technology activity of Hebei province operated low-level efficiency during the examined period. Therefore, it is necessary to improve the efficiency of science and technology activity, so that the limited resources of the region can be made the most used to satisfy the regional construction of innovative requirements.

References

1. Kao, C., Hwang, S.-N.: Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. European Journal of Operational Research 185, 418–429 (2008)
2. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30, 1078–1092 (2009)
3. Chen, Y., Zhu, J.: Measuring information technologies indirect impact on firm performance. Information Technology & Management Journal 5, 9–22 (2004)

4. Lindebo, E., Hoff, A., Vestergaard, N.: Revenue-based capacity utilisation measures and decomposition: The case of Danish North Sea trawlers. *European Journal of Operational Research* 180, 215–227 (2007)
5. Acs, Z.J., Luc, A., Attila, V.: Patents and innovation counts as measures of regional production of new knowledge. *Research Policy* 31, 1069–1085 (2002)
6. Niosi, J., Bas, T.G.: Biotechnology Megacentres: Montreal and Toronto Regional Systems of Innovation. *European Planning Studies* 11, 789–804 (2009)
7. Doloreux, D.: What We Should Know about Regional Systems of Innovation. *Technology in Society* 24, 243–263 (2006)
8. Asheim, B.T., Isaksen, A.: Regional Innovation Systems: The Integration of Local "Sticky" and Global" Ubiquitous" Knowledge. *The Journal of Technology Transfer* 27, 77–86 (2009)
9. Autio, E.: Evaluation of RTD in Regional Systems of Innovation. *European Planning Studies* 6, 131–140 (2005)
10. Ruggiero, J.: Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research* 111, 461–469 (2008)

Modeling and Verifying the Kernel of RTOS

ShengWen Gong

Qingdao University of Science & Technology, SongLing Rd. 99, 266071 Qingdao, China
qd_gsw@163.com

Abstract. This paper presents a study on modeling and verifying the kernel of Real-Time Operating Systems (RTOS). Advances in formally verifying such an RTOS both by refinement and by model checking approaches will be shown in this paper. The focus of the paper will be on verifying FreeRTOS. A number of ways to verify this operating system are investigated in the paper. At last, a preliminary set-up of verifying FreeRTOS using model checking is presented.

Keywords: model checking, real-time operating system, refinement, freeRTOS.

1 Introduction

Software should meet all of its requirements and fully conform to its specification. Testing and other quality assessment techniques can only help to approach this goal [1]. Proving that a piece of software will always work according to its specification requires the use of formal methods. Two approaches to using formal methods exist: correctness by construction [2] and post facto.

The first approach can be used to construct an implementation starting from an abstract formal specification. This technique is sometimes referred to as refinement and relies upon the use of theorem provers. The second approach, often called verification, extracts a model from existing code and uses state based technologies like model checking to determine the correctness of the model.

Verifying a real-time system (RTOS) is considered difficult for many reasons. A RTOS creates and operates on a number of concurrent user tasks, promises to schedule those tasks within guaranteed time-limits using some priority mechanism, guarantees non-interference between these tasks and between the tasks and the kernel, allows these tasks to communicate using queues, synchronizes these tasks using mutexes or semaphores, and interacts directly with hardware through device drivers, clocks and interrupt mechanisms.

Concurrent programming is an important technique in modern software development. Concurrency can improve computational efficiency and resource utilization. However, concurrent programs behave differently than sequential programs. Multiple executions of a concurrent program with the same input may exercise different sequences of synchronization events and produce different results. This non-deterministic behavior makes concurrent programs notoriously difficult to test and verify.

We do not target pervasive verification of the whole system stack consisting of hardware, compilers and the kernel but concentrates on verification of the kernel itself. This paper also does not target timing overhead verification of particular operating system functions such as context switching time. Such verifications are too much dependent upon the particular hardware used. Instead, the paper presents an overview of the issues involved in verifying a real-time kernel, discusses the levels of verification technology available and tries to set up a course to do formal verification using model checking.

2 Verification Methods

Before embarking upon a detailed discussion on methods usable for verification of a real-time operating system, a sketch of some properties to verify are given. It should be emphasized that only checking a real-time operating system for functional correctness is considered insufficient, safety properties and liveness properties should be specified and verified as well. Safety properties (reachability properties) are used to prove the absence of bad events, liveness properties show that a good event will happen eventually. In the sequel a discussion of various methods used for formal verification will be given.

2.1 Refinement Methods

Using the refinement approach, first an abstract specification of an RTOS is created. Then, a refined version (the concrete specification) is created. The two are related by a so-called abstraction relation. Proofs are needed to show that the concrete specification correctly represents the abstract one. Multiple refinement steps toward an executable language may be needed. Various tools are available to prove the system correct. Generally speaking, these proofs are performed by automatic theorem provers but steering these proofs requires considerable interactive user expertise.

Refinement technology enables the user to refine operations and data structures into more detailed ones. After a refinement step has produced a stateful sub-system, tools checking state properties such as model checkers may be used to prove relevant temporal behavior of the sub-system. A tool set having an integrated approach in this area is ProB [3].

However, refinements are usually done on individual operations and data structures. This requires reiterating the checking of temporal specifications after each refinement step. Little seems to be known about refinements having introduction rules spawning or refining temporally correct sub-systems.

2.2 State Based Methods

Contrasting with refinement based approaches, state based approaches (such as model checking) usually start from an existing implementation. Properties are being proved

over the implementation. Such properties can be proved using classical model checkers.

Languages for model checking allow the programmer to dynamically create processes; for each process an automaton is constructed. During model checking, the model checker will probe all possible interleavings between the processes exposing implementation errors that cannot be found using regular testing approaches.

To enable proofs of required properties, such a modeling language is usually quite restricted in its data and control structures. Properties to be proved are given in a temporal logic (propositional logic, predicate logic, Linear Temporal Logic (LTL), Computation Tree Logic (CTL), Timed-CTL (TCLT)). The model checker combines the automaton and the property specification and provides either a proof that the system indeed conforms to the specification or a counterexample showing a path leading to the error.

The model checker combines the automata describing the processes into one automaton; the size of the resulting automaton being the product of the sizes of the constituting automata; this is referred to as state explosion.

2.3 Abstraction Based Approaches

Performing model checking directly on code looks like a interesting proposal. Unfortunately, the state space resulting from a direct translation of code into an automaton is likely to explode already on small examples due to recursion and pointer constructs. Various authors have tried to remedy that situation by restricting the search procedures and the power of the query language. Usually, these systems have a limited search depth and use heuristics to search only in areas where the programmer expects an error. Unfortunately, in software verification, error traces are often quite long and therefore a large bound is needed; otherwise errors may be missed.

One way to reduce the size of the verification model is by applying the technique of predicate abstraction. This so-called CEGAR (Counter Example Guided Abstraction Refinement) method was introduced by Clarke et al. [4].

Predicate abstraction abstracts the data by keeping track of certain predicates on the data, leaving the control flow of the code unchanged. Each predicate in the concrete program is represented by a Boolean variable in the abstract program; the original data are eliminated. The abstract model and the (reachability) property are passed onto the model checker. If the property is proved correct on the abstract model, it is also correct on the concrete model. The abstract model has more behaviors than the concrete model. Therefore, a property shown to be false on the abstract model may be a spurious counterexample. A number of refinement steps may be needed to resolve the issue. In these approaches however, the query language is limited, e.g. only assertions and deadlock detection are supported and the expressivity of the programs is often reduced by disallowing e.g. threading and/or mutexes.

2.4 Translation Based Approaches

As checking the C language gives rise to many problems, several authors have proposed translations to various, more abstract, languages.

Translation to Java

As the C-language and the Java-language have much in common, translating the kernel into Java seems a viable option. The resulting Java code can then be tested using the Java Pathfinder system [5]. Difficulties regarding the translation of pointers and heap structures seem to disappear except for those places where void pointers are used to simulate genericity. However, mapping the concurrency mechanism of the RTOS to the Java thread model seems less intuitive. Inside JPF, the Java code is translated to code for a Virtual Machine enhanced for model checking. This VM only allows checking for invariants and deadlock to be performed on the Java program. But there is no support for temporal specifications.

Translation to Uppaal

The Uppaal model checker [6] has been designed to enable model checking of timed systems. The model checking algorithm is based upon TCTL, a Timed version of CTL (Computation Tree Logic) [7]. Although being one of the few model checkers being designed for timed systems, Uppaal seems to be incapable of checking the code of an RTOS because a Uppaal program consists of state machines and communication channels.

Translation to Promela

Spin has a restricted facility to include Ccode into a Promela program. Holzmann, Groce and Joshi [8] redesign that mechanism. They distinguish between matched data that is stored both on the search stack and in the state space of the model checker and unmatched data that is only stored on the stack. A large reduction in the amount of state space needed is obtained. However, the set-up influences the architecture of an application by placing state information in a contiguous area of memory.

Zaks and Joshi [9] describe another redesign of the same mechanism. Starting at multi-threaded C code, a typed byte code for a virtual machine is generated. The Spin system controls the virtual machine, allowing most proving mechanisms provided by Spin to be used. The paper mainly discusses extending the partial order mechanism needed. The largest example given is a 2800 line C program implementing a multi threaded communication system.

Gallardo et al. [10] describe an interesting approach using a two dimensional logic that combines time (control flow) and space (dynamic data structures). Control flow is specified using an extension of LTL; data is specified using CTL. Using the extended logic, they use Spin to model check properties of C programs, including dynamic allocation. By extending the checking of invariants in LTL, they are able to detect acceptance cycles. Unfortunately, not all temporal specifications can be tested and the examples given are rather limited in size.

3 Some Verified RTOS

3.1 What Is FreeRTOS

FreeRTOS is a portable, open source, royalty free,mini Real Time Kernel. It is a free to download and free to deploy RTOS that can be used in commercial applications

without any requirement to expose your proprietary source code. It has been downloaded more than 77,500 times during 2008. FreeRTOS is the cross platform de facto standard for embedded micro controllers.

The kernel of FreeRTOS has been written in C (2500 lines of .c-files, 800 lines of .h-files and some 100 lines of assembler). These are real lines of code; not comments. FreeRTOS has currently been ported to 24 different microprocessor architectures; ranging from ‘minimal’ 8-bit architectures to full blown 32-bit architectures. Depending upon the architecture, the kernel is compilable on 14 different compilers. Depending upon the application at hand, the user can include or exclude facilities like mutexes, tracing and scheduling mechanisms.

The source code is heavily parameterized upon the above items using the include mechanism of the C-language which makes extracting a proper subsystem to be used for verification a bit cumbersome.

In order to better understand the complexity of full verification, a number of features of the FreeRTOS kernel is given below:

Multiple tasks with priorities exist; tasks may be created in run-time. Coroutines with priorities are optionally available too. The scheduler uses prioritized preemptive or cooperative mechanisms to schedule those tasks. Tasks are synchronized by binary and counting semaphores and mutexes.

Lists of ready/suspended/blocked tasks involve heap storage, pointers and type casts. Dynamic memory management uses malloc() and free(). User definable communication queues allow for thread-safe FIFOs.

The stack-size has been preconfigured for each task.

The kernel controls hardware registers, timers and listens to interrupts.

The FreeRTOS documentation also defines an API programmers may use to invoke the functionality provided by the kernel. The kernel seems to have been originally developed for a small system not supporting memory protection and is therefore an example of a previous generation of real-time operating systems. Memory protection has been added recently for the new Cortex-M3 processors but will not be discussed here.

3.2 Some Commercial RTOS

Apart from the safe companion to FreeRTOS, various other systems claim to have been formally verified. Presumably the list below is far from complete, but it is assumed that the systems mentioned will give a good impression of what is available commercially.

WindRiver and QNX

These OSes are examples of so-called micro kernels. They have been ported to many architectures and are being used in millions of real-time systems. They have been certified according to many standards like IEC 61508, FDA 510K, ISO/IEC 15408 and DO-178B to various levels. Many of these standards are process level standards however; they prescribe the process a developer should follow to create dependable software. Very few of these standards require the use of formal methods.

OpenComRTOS

OpenComRTOS is advertised as a network centric RTOS with support for heterogeneous multiprocessors systems and multi core systems. The system has been ported to many architectures. Most of the C kernel code is generated and checked according to MISRA. The kernel code is reportedly much smaller than the competitors' code. OpenComRTOS is suited for level 3 or 4 SIL applications. Full source code and formal application reports are available on special contract. The system has been partially verified using the TLA+/TLC system. TLA/TLC was used primarily as an architectural design tool. The TLC model checker declares every action as a critical section, disallowing the verification of concurrency present in the system. This resulted in an avoidance of global data structures in the architecture.

PikeOS

The PikeOS is a further development of the L4 micro kernel. The PikeOS web-site states that the kernel is certifiable to safety-critical standards (DO-178B, IEC 61508 and EN 50128). The authors have used the Microsoft Verifying C Compiler which allows them to formulate contracts and invariants stored as annotations within the source code. So far, they seem to concentrate on the system call level; an example of a system call changing the priority of a thread is given. An interesting feature is the verification of assembly code and hardware semantics. It remains a bit unclear from the paper how the approach would scale up towards checking the whole of the kernel.

4 Verifying FreeRTOS

Based upon all the material presented herein, it remains to present a suitable proposal for verifying the kernel of FreeRTOS. Basically, a choice can be made between the refinement based approaches and the state based approaches.

Regarding the first category it can be said that, as exemplified in this paper, much successful work has been done already. In the category of state based systems, the CBMC-like approaches seem unsuited to handle the verification of an OS because these approaches have difficulties handling threads, concurrency and loop unrolling problems. An operating system basically consists of an infinite loop and in the case of FreeRTOS even user tasks must be programmed as infinite loops. In addition, CBMC-like systems are usually limited to proving safety properties.

The SLAM approach requires special mentioning here. SLAM targets correctness of device drivers and in particular the temporal aspects of ordering of API-calls to a device driver. Some of the work of SLAM is valuable to the FreeRTOS-work, but the API-part seems less applicable here because FreeRTOS only has a single API-layer directly accessible from application programs. Checking such APIs involves checking user programs. A test harness containing such programs is currently unavailable but developing the SLAM technology for users of the FreeRTOS would be very fruitful. Such a test harness would however give more credentials to the correctness of the user program than to the correctness of the FreeRTOS itself.

Testing temporal specifications of an RTOS using a model checker as e.g. Spin suggests the following two routes:

The matched data/unmatched data approach by Holzmann et al
 A hand translation from C code into Promela code

4.1 The Holzmann/Zaks Approach

In this approach some problems are to be expected, e.g.:

The dependency upon the pthreads library must be replaced by the more general mechanism of interrupts and context switching.

Using this method, many probes must be inserted in the code. Finding a systematic method to find and insert probes is much needed.

Solving the problem of limited concurrency.

4.2 Translation by Hand

When C-code is hand-translated into Promela, several abstraction issues need resolution:

When verifying FreeRTOS there seems no need to solve the general C to Promela translation problem and there is no need to solve the general pointer problem. In the FreeRTOS kernel, pointers are mainly used to construct linked lists of a few data types only. The storage model for the heap may be simulated with an array construct; pointers are replaced by indexes. This will lead to a faithful transformation; C-code can be translated by hand in a one-to-one way into Promela. Preliminary experiments show a large state vector however. This is to be solved by using a special compression mechanism in addition to the mechanisms built into Spin already. If state compression proves insufficiently effective, a less faithful model without indexes can be constructed.

In fact, the kernel and all application programs form a sequential piece of code. Context switching only provides an illusion of parallelism. Real concurrency stems from I/O devices and interrupts. I/O and interrupts can be implemented using a very small state space, thus minimizing state explosion. As a further reduction, limiting the amount of context switches can be tried.

One particular problem is the difference in atomicity between Spin and machine instructions. In Spin, every statement is, by definition, atomic. With machine code, interrupts can occur on every machine instruction; sometimes even inside a machine instruction. Because the Promela code can be interrupted before and after every instruction, sufficient interruptible locations are present and no difficulties are foreseen.

In FreeRTOS, a context switch will save all registers on the stack and later restore the previous context. As registers are unavailable in a Spin simulation, it seems sufficient to reduce stack handling to handling return addresses.

The number of Spin data types is quite restricted compared to the number of C data types. E.g.: C knows about the data types char, signed char and unsigned char. Promela only has a byte data type which is equivalent with unsigned char. Promela has no long data type either.

The above approach has been used on a small scale in a web report on the formalization of the Minix kernel.

Currently, both methods mentioned above seem to be the most fruitful ways forward and both implementations are at this moment being worked on.

5 Summary

This paper has provided an overview on various methods to increase the confidence in the functional and temporal correctness of a real-time operating system. In particular, proving safety and liveness of such an operating system is considered very important. Both refinement based and state based approaches have been discussed. In the last, two promising techniques, the Holzmann/Zaks approach and the direct translation of C to Spin have been selected for FreeRTOS investigation.

References

1. Dijkstra, E.W.: Notes on structured programming. In: Dahl, O.J., Dijkstra, E.W., Hoare, C.A.R. (eds.) *Structured Programming*. Academic Press (1972)
2. Jones, C., O'Hearn, P., Woodcock, J.: Verified software: A grand challenge. *IEEE Computer: Software Technologies* 39, 93–95 (2006)
3. proB, <http://www.stups.uni-duesseldorf.de/proB>
4. Clarke, E., Grumberg, O., Jha, S., Lu, Y., Veith, H.: Counterexample-guided abstraction refinement for symbolic model checking. *Journal of the ACM* 50 (2003)
5. Java Path Finder, <http://babelfish.arc.nasa.gov/trac/jpf>
6. Uppaal. The Uppaal model checker, <http://www.uppaal.org>
7. Baier, C., Katoen, J.P.: *Principles of Model Checking*. The MIT Press (2010)
8. Holzmann, G.J., Joshi, R., Groce, A.: Model driven code checking. *Autom. Softw. Eng.*, 283–297 (2008)
9. Zaks, A., Joshi, R.: Verifying Multi-threaded C Programs with SPIN. In: Havelund, K., Majumdar, R. (eds.) *SPIN 2008. LNCS*, vol. 5156, pp. 325–342. Springer, Heidelberg (2008)
10. Gallardo, M.M., Merino, P., Sanán, D.: Model checking dynamic memory allocation in operating systems. *J. Autom. Reasoning* 42, 229–264 (2009)

Improved Party Building Based on Information Technology

Qiangqing Wang¹ and Jierui Wang²

¹ HeBei United University, 063009 TangShan, China

² KaiLuan (Group) Co., Ltd. Hospital, 063000 TangShan, China
feiyue063000@163.com

Abstract. What informationization is that basing on the widespread availability of modern information technology, people significantly improve the functionality and efficiency of the various social activities by improving the management of information resources and utilization levels. One of the important methods to strengthen ruling capability of Party and educate Party members is using network information technology to develop grass-roots party building work. Therefore, it is essential to study on the significance and restrictive factors of the work of Party building information technology, to give more proposal on how to do it well, by which the application of information technology areas and the efficiency of work of party building can be impressively promoted.

Keywords: Information Technology, the Party Building, Promote.

1 Introduction

Into the 21st century the era of network information, e-party construction is to strengthen the party's ruling capacity building, and an important way to improve the party's leadership and governance. Comrade Jiang Zemin pointed out:"due to the development of the information network, has formed a new ideological and cultural positions and ideological and political struggle positions our Party building work, ideological and political work, organization, and should adapt to the information network characteristics." General Secretary Hu Jintao pointed out that "the use of information network technology to strengthen and improve Party building, improve the efficiency of the work of party building." The use of network information technology to carry out grassroots party building is to strengthen the governing capacity building of grassroots party committees, and an important means of management education of Party members.

2 The Necessity and Significance of a Work of Party Building Information

Information Management information refers to the basis of the wide spread of modern information technology, information resources management and utilization levels, substantially improve the functionality and efficiency of the various social activities.

The work of party building information refers to the basis of the wide spread of modern information technology, the work of party building of the ways and means of the breakthrough and change with the promotion of social development and information technology, by improving information resources management and utilization levels, party building organizations a substantial increase in the functionality and efficiency of the work. Closely integrate the work of party building and information construction is an effective means to achieve the party building management innovation, technological innovation.

2.1 The Information Technology Requirements of the Work of Party Building Innovation

The rapid development of information networks, not only on the economic, political, military, cultural and other fields with a significant impact on production and lifestyle of the people of more profound change at an unprecedented impact and penetration, and vigorously promote the development of information technology, has become an important and urgent strategic task of the new stage of China's economic and social development, which requires the work of party building is going to follow the development of the times. Ability to seize the opportunity to use IT and the Internet, building a network of party building positions, to improve the party's ability to strengthen the online public opinion to guide, to maintain the Party's advanced nature is the work of party building facing severe challenges and tests. Political life of the information network has become an inevitable trend of social development, the work of party building in the major trend in the information society must continue to play to our strengths and gain the initiative, it must keep pace with the times, as soon as possible to achieve integration with modern IT standards, improve their own scientific and technical content, truly the times.

The party's seventh session of the Fourth Plenary Session of the first clearly to promote grass-roots party organizations, require the use of the Internet, cell phones and other media innovation means of party work, opening up the party organization space, so that the party's grassroots organizations always and social development synchronization. Construction of the work of party building information covering various aspects of the work of party building, and reflects the CPC Central Committee attaches great importance to this work, in close connection with the work of party building and the construction of information technology, and continuously improve the scientific level of the work of party building, both the work of party building management innovation and means of technological innovation, the new period to strengthen the Party's governance capability and maintain the Party's advanced nature of the inevitable requirement.

2.2 Information Is a New Field of Work of Party Building

Party building information technology with the government information technology and e-government development, function and role is bound to be gradually assuming the role, content of services involved in the ideological and theoretical propaganda,

organizing and coordinating the exchange, convey system files, daily transaction processing, management of Party membership, education and training, datastatistics and other aspects.

High penetration of information technology has provided a guarantee for the work of party building. Information technology training, the development of intelligent computer-based tools for the new productivity and benefit in the process of social history. The progress of modern science and technology has provided the material assurance and technical support for the modernization of the ideological and political work. In ideological and political aspects of organization building, and democratic supervision, information technology has the characteristics of the flat extension and sharing of information resources, highly developed information technology today, we can use a variety of media, and extensive collection of ideological and political information to carry out ideological education; can take advantage of the network of party management, office automation, the work of research. Party building information technology channels for inner-party democracy and party-communication channels to broaden and smooth, and is conducive to the expansion of the party members and people's right to know and the direct supervision of the party's activities, to compensate for the shortcomings of the traditional vertical management system.

High-speed information technology to create the conditions for party building work carried out efficiently. Information technology to the dissemination of information breaks through the limitations of print media at the time, space, depth, and become the source of the spread of an all-weather, global, highly efficient, agile, rapid response characteristics. To deal with all aspects of risk and emergencies, information networks on the transmission of information advantage especially, is conducive to the rapid dissemination of the work of party building decision making and implement, but also conducive to resolve conflicts, consolidating the party's position and guide the work of party building how to strengthen the modern crisis management, to do this promptly and properly respond to emergencies.

Information on the diversity of the work of party building a new carrier. Since the mid-1990s, the positive role of information technology in the ideological and political work began attracting widespread attention. "Internet Party" has become an important working tool for all levels of party organizations and party workers. Can establish the position of the party in the field of ideology on the Internet network to carry out Party building site construction, timely and accurate dissemination of the work of party building, the Party organizations at all levels, Party members and the majority of Internet users directly from the Internet at any time to obtain the necessary party's knowledge and information to establish the correct guidance of public opinion, and use audio and video, graphics, animation and other multimedia forms to show the main theme of the work of party building in the new period. To carry out the construction of electronic party to party work process standardization, decision-making process is open, the democratization of management functions, and enhance the efficiency of the work of party building, service and implementation of scientific organizations at all levels to provide timely, accurate and comprehensive information support for democratic decision-making services. To carry out the construction of online Party schools, cadre training colleges and institutions of higher education and

resource sharing, and inherit the traditional party school education at the same time, the Online Party to do the new era the whole party to learn the theory of Marxism and socialism with Chinese characteristics classroom.

3 Constraints of the Current Work of Party Building Development

Information technology development in the context of contemporary society, the factors of influence and restrict the party building the organization of the process of information are the following:

3.1 Party Building Information Technology Hardware Limited

In accordance with the construction of "saving, efficient, clean the request of the Government departments at all levels has increased with the additions of computer hardware such as office facilities, the basic structures of the" paperless office "and" office automation "platform. But from a practical point of view, the input of the hardware seems limited, there is a gap, resulting in the work of party building information technology subject to certain restrictions on the true meaning of "office automation".

3.2 Information Tools Utilization Is Not High

In the daily work and life, we can find some phenomena, such as: individual work time playing games, working hours online to read some light readings, causing the computer to become some people's "game" and "reading". There is not in accordance with the requirements of the proper use of, leading to severe loss, resulting in unnecessary waste. These are the tools utilization is not high performance.

3.3 Cadres Ability and Quality to Be Improved

On the one hand, the main leadership of some units and some cadres did not generate enough attention to see the necessity and inevitability of the party building the organization of information technology are not allowed to see through, to see far, and mixed the use of the skills of the cadre of information technology tools. On the other hand, everyone who comes into a compulsory system, some real technical personnel can not enter the circle. The overall capacity requirements there is still a gap between the education and training times to be further strengthened.

4 Do the Work of Party Building Recommendations

Party building information technology building is a huge project, which includes many aspects, and close ties with and between the work and complement each other,

strengthen party building information technology building must be Systems Analysis, the full implementation in order to ultimately implement.

4.1 To Further Improve the Understanding of the Information on the Work of Party Building

Awareness-raising is the premise to solve the problem, if you do not attach importance to the opportunities and challenges of information technology, it is possible to lose the initiative in the virtual world, then lost the leadership of the real world. We want to party building information included in the important task of party building work. Levels of party organizations, party workers should be further recognized that the party building information technology is an important part of Party Building and the party's governing capability, an innovation of the party's leadership in the information technology environment and governance. According to the new situation and new tasks and new requirements, and effectively improve the understanding of the use of information technology to carry out the importance of the work of party building.

4.2 To Further Improve the Party Building Information Technology Service Platform

With the continuous advance of the information process, the use of more and more new tools, more widely, to take advantage of these technology as the party building the organization of services, to ensure the precision technology into a fine business. First, daily work, organizational work is characterized by a wide range of large and heavy task to further strengthen the normative use of information equipment, good equipment, good platform, and continue to improve the daily work of the organization departments of high-quality, high volume and efficient. The second is further refined with specialized information management system, such as: file management system, cadre management system, party members repository, effectively make the best use through use fine to play its biggest role. In the audio-visual work, information and research work to further strengthen the use of information technology, eye-dotting ceremony for party building organizations.

4.3 To Further Strengthen Party Building Work in Information Technology Personnel Training

Talent is the key to career development, and should pay attention to the work of party building information technology and personnel training, set up a quality team of professionals to enhance the level of work of party building information. First, we will formulate a set of party building training methods and training systems, and the introduction of party building information technology educational content in the school's training system. Second, to study the development of party building information technology personnel management practices, in order to attract, motivate, and management professionals. Hiring a group familiar with the information work of

the party and government organs, the formation of the Party Building Information Construction Expert Advisory Group to provide advisory services for the construction of party building information technology, and actively make suggestions for party building information technology.

5 The End

Rapid development of information technology for national economic and social production of all areas of life had a profound impact on the building of the party should seriously explore the significance of IT in the work of party building, party building in the information technology platform, with the power of information technology to achieve the optimal reorganization of the party organization, management and workflow, so that the work of party building along with the times, to play a greater role in IT.

References

1. Chen, S., Wu, J.: Party Building Strategy Selection. School Party Building and Ideological Education 9, 26–27 (2008)
2. Wang, S.: Strengthening and Improving Party Building under Conditions of Informatization. HuXiang Forum 5, 62–66 (2011)
3. Hong, X.: Promote the Construction of E-Party Problems and Countermeasures. ShangHai Party and Party Building 7, 44–46 (2008)
4. Chen, X.: Implementation of the Party Building Management Practice and Thinking. China Science and Technology Information 11, 36–39 (2006)
5. Tang, J., Song, Y.: University Network Characteristics of the Work of Party Building. Current Situation and Strategic Choices. Theory and Practice of Contemporary Education 12, 55–58 (2011)
6. Li, X., An, H.: Research on the University Computer System. Coal Technology 8, 254–255 (2011)
7. Huang, R.: Application of Computer Technology and Network Technique in Computer Room Management. Value Engineering 1, 178 (2011)
8. Zhang, Y.: The Application of Network Technique in the University Administrative Management. Computer Knowledge and Technology 6, 826–827 (2010)
9. Zou, X.: Internet Age Party Building in Universities. School Party Building and Ideological Education 7, 35–36 (2009)
10. Gu, L., Wang, K.: Overview of Modern Network Technology. Software 3, 11–13 (2011)
11. Feng, P., Pan, X.: By Network Party-Building to Promote the Innovation and Development of Party-Building Work in the Higher Vocational Colleges. Journal of WuXi Institute of Technology 10, 27–29 (2011)

Mobile Customer Satisfaction Evaluating System

Yan Gao^{1,2}, Lihui Zhou¹, and Linfan Li³

¹ College of Science, Hebei United University, Tangshan, Hebei, 063009

² Center for Quantitative Economics of Jilin University, School of Business,
Changchun 130012, Jilin Province, China

³ People's Armed College, Hebei University, No. 42 Ping'an South Avenue,
Shijiazhuang 050061, Hebei Province, China

Abstract. For mobile communication enterprises, improving Customer Satisfaction is benefit to transformation of the business strategies and is useful for them mastering users' requirement and expectation. It is even more available in improving the ability of gaining profit, so it is necessary to analyze Customer Satisfaction. In this article, we research customer satisfaction for the Mobile Corporation, and propose the final solving project. We determine the weight for each index by AHP method, and then we will calculate the scores of the Customer Satisfaction. At last, we will give the strategy and the advice for the problem in improving customer satisfaction.

Keywords: Customer Satisfaction, Evaluating System, AHP.

1 Introduction

Definition of customer satisfaction is given out by many literatures, the most common one is put forward by Philip Kotler, he thinks that satisfaction prefer to the feeling attitude of happy or disappointment after comparing the feeling effect to his expectation. The level of satisfaction is the difference function between the effect which can be felt and the expectation. Degree of satisfaction usually means the satisfaction degree of the products or service from an enterprise.

The characters of mobile communication industry decides on its industry property, in modern society, it is getting more and more intense along with the market competition environment and the expectation of the customer, in order to develop well, mobile operators must consider some important problems, such as how to make sure the number of customers in line stable, how to develop new customers, how to develop data business and how to set up good enterprise image. The basic method to solve the problems above is to know the satisfaction of the customers by diagnoses, then show the problems by satisfaction analysis, at last, propose the way to deal with the keeping the enterprise developing well. In this article, we will take some province mobile company for example to analysis the satisfaction degree.

2 Plans in Study of Mobile Customer Satisfaction

At first, some basic concepts will be put forward: The competitors of mobile are China Telecom Corp. and China Unicom Ltd., the Strong competitor is defined as the one who get the higher score for satisfaction, Leading means the D-value of mobile data and the strong competitor data, from leading value, we can know the compare of relative indexes of China Mobile, China Telecommunications and Chinese Unicom, so that it will supply Operation Reference for our results.

Second, we classify the objects we study into China Mobile customers, China Telecommunications customers and Chinese Unicom customers according to the market conditions; then we classify the China Mobile customers by brand: GSM, EASYOWN and M-zone, and we draw 120 examples from GSM, 180 from EASYOWN and M-zone, we separately draw 80 from China Telecommunications and Chinese Unicom.

We want to know the mobile communication service requirement for customers, measuring the satisfaction, leading degree and the performance in main business course, so that we can find problems, find the effective method to improve the customer satisfaction.

Around the problems which customers concern most, we will define the problem in ten aspects: telephone charge, reward points, payment, advertise, price, new business, business hall, network quality, sales promotion and customer service hotline. We collection data by CATI, then study the model.

The system of Index Evaluation in divide into three levels: 1) degree of satisfaction, 2) telephone charge, reward points, payment, advertise, price, new business, business hall, network quality, sales promotion and customer service hotline; 3) the details of the index in second level, for example, the new business can be divided into: if the open is confirmed, if the cost is reasonable, if the cancel is of convenience, if the business is easy to use.

In analysis, we use qualitative analysis to dispose the information of no data. We use quantitative analysis to show the data in table. At last, we will get the result of satisfaction, and then we will propose the suggestion for improving satisfaction.

3 Customer Satisfaction Analysis

3.1 Computing Method of Customer Satisfaction

Satisfaction of different levels is the weighted average of the next index, we denote CS as customer satisfaction whose formula is shown below:

$$CS = \sum_{i=1}^n W_i X_i \quad (1)$$

In which, W_i is weight, X_i is the i th value of satisfaction for the next indices, n denotes number of the indices.

The formula in (1) can be described as:

Value of satisfaction for whole mobile customers

$$= \sum \text{scores of second layer index} \times \text{weight} \quad (2)$$

$$\text{Scores of second layer index} = \sum \text{score of third layer index} \times \text{weight} \quad (3)$$

For the third layer, the computing formula is:

Scores of third layer index

$$= \text{Sum of scores for all questionnaire}/\text{number of questionnaire}. \quad (4)$$

3.2 Determination of Weights for Each Level

We use the Analytic Hierarchy Process (AHP) method to determine the weights for each layer. AHP is a systematic, hierarchical method which is combined with qualitative and quantitative analysis. First of all, we should put the problems into different levels, according to the characters and target, we decompose the problems into target layer, criterion layer and project layer, then compare the influence from different project to target, and give out each weight for every project. Kernel of AHP is to determine the influence from criterion layer to target layer and the effect of each project to every criterion (weight) by Pairwise Comparison Matrix , in application, we can determine the importance degree according to 1-9 scaling method(shown in Table 1),which is proposed by TL Satty.

Table 1. 1-9 scaling method for AHP

Relative importance	Score	Explain
Equally important	1	i and j are equally important
Slightly important	3	i is slightly more important than j
Quite important	5	i is more important than j
significant important	7	i is significant more important than j
Absolutely important	9	I is absolutely important than j
Between the two degree	2,4,6,8	I and j Between the two degree

Set index i get score r_{ij} relative to j , then index j get score $r_{ij} = 1/r_{ji}$ relative to i , according to this criterion, we establish Pairwise Comparison Matrix R ,

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (5)$$

This Matrix is positive reciprocal matrix, according to AHP, $RW = \lambda_{\max} W$, and $\xi W = 1$. In which, λ_{\max} is the maximum eigenvalue, ξ is constant vector,

$\xi = (1, 1, \dots, 1)$, $W = (W_1, W_2, \dots, W_n)$ is the uniformed Eigenvector corresponding to maximum eigenvalue for R , we usually use the each component of W as weight of each index.

3.3 Statistical Analysis for the Satisfaction of the Second Layer

In satisfaction assessment system, the second layer denotes the satisfaction of some business process. First we study the satisfaction to new business. We can use formula (4) to get score of the third layer. Then we will get scores for the each detail items of new business, the results are shown below:

Table 2. Satisfaction Score New Business Detail Items for Each Brand

brand	The new business items						If cost Reasonable
	If business meet demand	If transparent clear	If Easy to use	if convenient	Cancel		
GSM	85.12	88.45	87.23	85.63	77.24		
M-zone	84.14	87.16	85.17	86.24	76.92		
EASYOWN	86.84	88.95	88.26	88.98	79.87		

Data sources: survey results obtained

We get weighted average of each item per brand, the weight is determined by effect from the items to satisfaction, we use AHP, Pairwise Comparison Matrix is given below:

$$R_{new} = \begin{bmatrix} 1 & 1/3 & 5 & 4 & 3 \\ 3 & 1 & 6 & 5 & 4 \\ 1/5 & 1/6 & 1 & 1/3 & 1/5 \\ 1/4 & 1/5 & 3 & 1 & 1/4 \\ 1/3 & 1/4 & 5 & 4 & 1 \end{bmatrix} \quad (6)$$

The uniformed Eigenvector corresponding to maximum eigenvalue is, $W_{new} = (0.260, 0.464, 0.042, 0.074, 0.160)$, so the weight is $(0.260, 0.464, 0.042, 0.074, 0.160)$, so that we can calculate the satisfaction of new business per brand, the values are 85.524 point, 84.578 point and 86.916 point, then we get the weighted average of the three brands, According to the contribution of three brands to income, we determine the weights subjectively as: GSM 25%, M-zone 20%, EASYOWN 55%, at last, we calculate the satisfaction score of new business by weight, the score is 85.69 point. We also get the satisfaction score of new business for China Telecommunications and Chinese Unicom, score of new business for China mobile is 3.89 point higher than strong competitor.

We can also get the scores of satisfaction for other business processes by AHP method. Telephone charge get 88.53 points, reward points 80.59 points, payment 86.21 points, business hall 84.40 points, network quality 84.26 points, sales

promotion 83.18, . and customer service hotline 90.46, all lead to strong competitor. In which, the leading degree of sales promotion is the least, the score of hotline is the highest.

3.4 Satisfaction Analysis for Indices of the First Layer

We can use AHP method to analysis the satisfaction for customers in the whole province. First, we give out the Comparison Matrix,

$$R = \begin{bmatrix} 1 & 1/2 & 3 & 5 & 1/4 & 1/2 & 1/3 & 1/5 & 4 & 1/3 \\ 2 & 1 & 3 & 5 & 1/5 & 3 & 1/2 & 1/5 & 4 & 1/2 \\ 1/3 & 1/3 & 1 & 3 & 1/5 & 1/3 & 1/5 & 1/7 & 1/2 & 1/5 \\ 1/5 & 1/5 & 1/3 & 1 & 1/7 & 1/4 & 1/6 & 1/7 & 1/3 & 1/5 \\ 4 & 5 & 5 & 7 & 1 & 5 & 3 & 1/3 & 5 & 4 \\ 1/2 & 1/3 & 3 & 4 & 1/5 & 1 & 1/4 & 1/6 & 1/2 & 1/4 \\ 3 & 2 & 5 & 6 & 1/3 & 4 & 1 & 1/3 & 4 & 2 \\ 5 & 5 & 7 & 7 & 3 & 6 & 3 & 1 & 6 & 4 \\ 1/4 & 1/4 & 2 & 3 & 1/5 & 2 & 1/4 & 1/6 & 1 & 1/4 \\ 3 & 3 & 5 & 5 & 1/4 & 4 & 1/2 & 1/4 & 4 & 1 \end{bmatrix} \quad (8)$$

The uniformed Eigenvector corresponding to maximum eigenvalue is $W = (0.074, 0.088, 0.033, 0.022, 0.181, 0.024, 0.177, 0.272, 0.014, 0.115)$, So the weight $(0.074, 0.088, 0.033, 0.022, 0.181, 0.024, 0.177, 0.272, 0.014, 0.115)$, the satisfaction score of overall province is 85.861, We also get the whole satisfaction score for China Telecommunications and Chinese Unicom, they are 79.35 points and 82.66 separately, leading value is 3.201 points.

4 Conclusion and Countermeasures

Although mobile customer satisfaction for the whole province and the satisfaction of each business process is better than strong competitor, there are still a lot to do in order to improve the satisfaction of customers.

For new business, customers unsatisfied with the transaction process, the main problems are open without notice and not convenient to cancel. Second, the new business is expensive; and fuzzy charging is the main third reason. In order to improve the customer satisfaction to new business, the mobile company should strengthen internal management. They should completely eradicate the behavior of open business without noticing customers. When they are advertising for the open way of new business, they should also tell customers how to cancel the business. At last, they should lessen the cost of customers by different methods.

For network quality, sometimes network signal is not stable is the main reason for customer's dissatisfaction. At present, network superiority is less and less for china

mobile, in some areas, the network is even not as good as competitor, they should strengthen the construction of base station, so that they can keep the leading role.

Concern of customers to business hall is waiting time and if seldom meet system fault, the index of if the waiting time can be accepted get low score, they should consider increasing staff and equipments in backstage, and increase mobile service staff, so that they can supply the customers waiting with problem solving and warm service.

For other business process, they should also concern to the items which get low scores, and find the problems. Then they need to find the solution for the problems, Whatever they do, they should consider from customers angle, they should have the same experience of customers, so that they can solve the problems with chariness and responsibility, so that the whole customer satisfaction will be Significantly improved.

Acknowledgments. Science research and development guide plan of Tang Shan (No.111102033b), (No. 111302013b) Nature Science Fund of Hebei province, (No.A2010000905, No. A 2012209030) Hebei Higher Social Science Research 2011 Annual Fund (No.SZ2011518); Science research fund of United University (No.z201114).

References

1. Peterson, R.A., Wilson, W.R.: Measuring customer satisfaction: fact and artifact. *Journal of the Academy of Marketing Science* 20, 61–71 (1992)
2. Tian, X.: Construction of Index Evaluation System for Customer Satisfaction of Commercial Banks based on AHP. *Consume Guide* (9), 69–70 (2007)
3. Bruce, W.M., Blackburn, J.W.: Blancing Job Satisfaction and Performance, pp. 20–45. Quorum Books, Westport (1992)
4. Oliver, G.L.: Effect of satisfaction and its antecedents on consumer preference and intention. *Advance in Consumer Research* (1981)
5. Luo, X., Homburg, C.: Neglected Outcomes of Customer Satisfaction. *Journal of Marketing* (4) (2007)
6. Yu, G.: Market survey and predict. China International Broadcasting Publishing House (2000)
7. Tang, S.: Telecom service quality and service marketing. Beijing University of Posts and Telecommunications Press (2002)
8. Cao, X.: Mathematical models in Management science. Peking University Press (2006)
9. Liu, J., Bai, D., Ji, X.: Modern enterprise should try to improve customers' satisfaction. *China Collective Economy (last ten-day)* (7), 39–40 (2007)
10. Zhai, Q.: Study about enterprise customers' satisfaction. *Productivity Research* (08), 108–109 (2007)
11. Li, B.: Analysis of enterprise customer satisfaction improvement method. *Consume Guide* (4), 8–9 (2007)

Performance Analysis of Transcritical CO₂ Sewage Source Heat Pump by Visual Basic Program

Hongli Wang¹, Huiqin Liu¹, and Jingrui Tian²

¹ College of Metallurgy and Energy, Hebei United University, Tangshan, 063009, China

² Library, Hebei United University, Tangshan, 063009, China

Abstract. Based on the problems of freon refrigerants ODP and GWP, the mathematical model of CO₂ transcritical sewage source heat pump cycle with a throttle valve and an expander were established. Based on Visual Basic program software, the performance analysis platforms were also developed. Regardless of the transcritical cycle with a throttle valve or an expander cycle, there is an optimum discharge pressure. The cycle with an expander has an efficiency of expander limitation value, which is the determination value of the throttle valve cycle and the expander cycle. With increasing of evaporating temperature, the performance of the two kinds cycles increase. With decreasing of outlet temperature of gas cooler, the performance of the two kinds cycles decrease. The compressor efficiency or the expander efficiency are higher, the corresponding system performance is higher. Under the same comparison conditions, the average cycle performance with an expander is better than with a throttle valve cycle. The fundamental data were obtained for improving cycle performance and developing the products of CO₂ heat pumps water heater.

Keywords: thermodynamics, CO₂ sewage source heat pump, throttle valve, expander, performance analysis.

1 Introduction

With the growing awareness of dual threats of global warming potential and ozone depletion potential [1], people have paid more attention to the substitute of CFCs and HCFCs. Re-employ natural refrigerants are a safety choice [2]. The development of environment-protective natural CO₂ substitution research work has the vital significance [3].

Large amounts of sewage discharged into the river in the city, causing severe water pollution, but also causing thermal pollution [4]. Because little temperature fluctuations and the volume stability of city sewage, which CO₂ heat pump[5] can be used to recover the large number of low-temperature waste heat in wastewater [6], also the efficiency of heat pump can be improved. On the one hand, the use of the conventional refrigerant can be reduced, thereby reducing the influence of ODP and GWP. On the other hand, the CO₂ refrigerant can be recycled to use in air conditioning and heat pumps. In addition, heat pumps are energy-efficient use of

clean energy in the form, which reduce the power consumption of the air conditioning and heat pump field.

At home and abroad, the analysis of CO₂ transcritical cycle [7] is adopted foreign software. Not only the fee is expensive, and intellectual property is protected, which is difficult to popularize the CO₂ heat pump technology in the country. Based on Visual Basic program software, the performance analysis platforms were also developed. The transcritical CO₂ heat pump cycle with a throttle valve [8] has high throttling losses, which the performance can be significantly improved instead of the throttle valve by expander [9]. The comparative analysis of cycle performance with a throttle valve and an expander were employed, and the fundamental data were obtained for improving cycle performance and developing the products of CO₂ heat pumps water heater.

2 System Composition and Thermodynamics Analysis of Transcritical CO₂ Sewage Source Heat Pump with a Throttle Valve and an Expander

2.1 System Composition of the Two Kinds of Cycles

A typical transcritical carbon dioxide refrigeration cycle consists of a compressor, a gas cooler, an evaporator and an expansion device, which in this study is either a throttling valve or an expander [10]. The schematic plan and T-s diagram of the CO₂ heat pump with a throttle valve and an expander are shown in Fig.1 and Fig.2, respectively. In the T-s diagram, the line 1-2_s-3-4_s-1 shows the ideal heat pump cycle with an expander, the 1-2-3-4-1 represents the actual heat pump cycle with an expander and the 1-2-3-4_h-1 shows the actual heat pump cycle with a throttle valve.

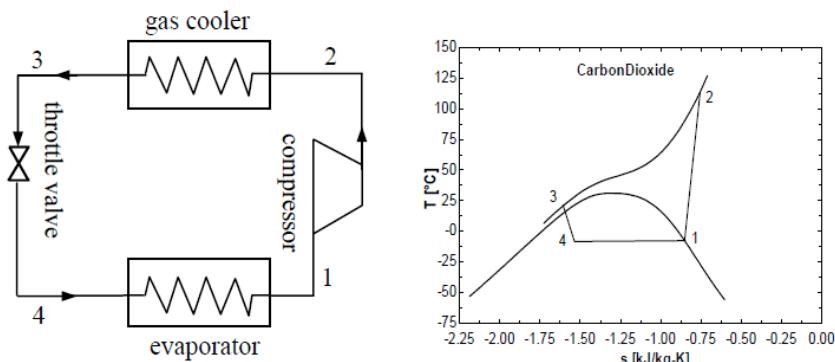


Fig. 1. Schematic plan and T-s diagram of the throttle valve Cycle

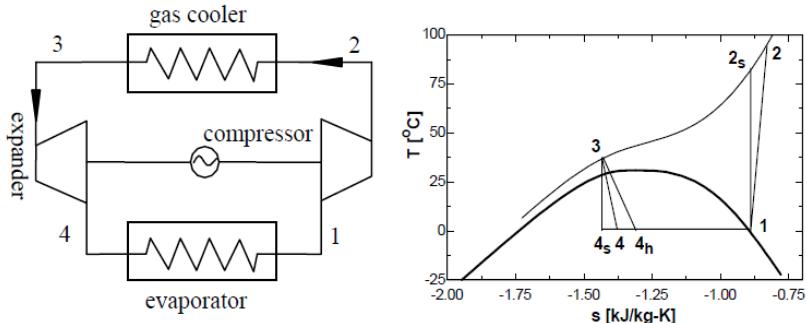


Fig. 2. Schematic plan and T-s diagram of the expander Cycle

2.2 Thermodynamic Analysis of the Two Kinds of Cycles

The coefficient of performance (COP_v) of transcritical CO_2 cycle with a throttle valve is defined as

$$COP_v = \frac{h_1 - h_4}{h_2 - h_1} \quad (1)$$

Where, h_1 , h_2 , and h_4 denote the specific enthalpies of CO_2 cycle with a throttle valve at the corresponding points, respectively, as shown in Fig. 1.

The coefficient of performance (COP_{exp}) of transcritical CO_2 cycle with an expander is defined as

$$COP_{\text{exp}} = \frac{h_1 - h_4}{(h_2 - h_1) - (h_3 - h_4)} \quad (2)$$

Where, h_1 , h_2 , and h_4 denote the specific enthalpies of CO_2 cycle with an expander at the corresponding points, respectively, as shown in Fig. 2.

The assumptions in the study are given as follows:

- (1) Compressor efficiency in practical and ideally is 0.75 and 1, respectively,
- (2) Expander efficiency in practical and ideally is 0.75 and 1, respectively,
- (3) Range of evaporating temperature varies from 5 to 10°C,
- (4) Range of outlet temperature of gas cooler varies from 28 to 38°C,
- (5) Saturated state at the evaporator outlet.

Based on the above calculation conditions, steady state operation, negligible pressure drop and heat loss of heat exchanger and other pipes are assumptive.

3 Development of Calculation Platform

In order to compare the performance of the two kinds of cycles, the performance analysis platform was designed in this paper by Visual Basic program software. The calculation platform of the two cycles are shown in Fig. 3 and Fig. 4, respectively.

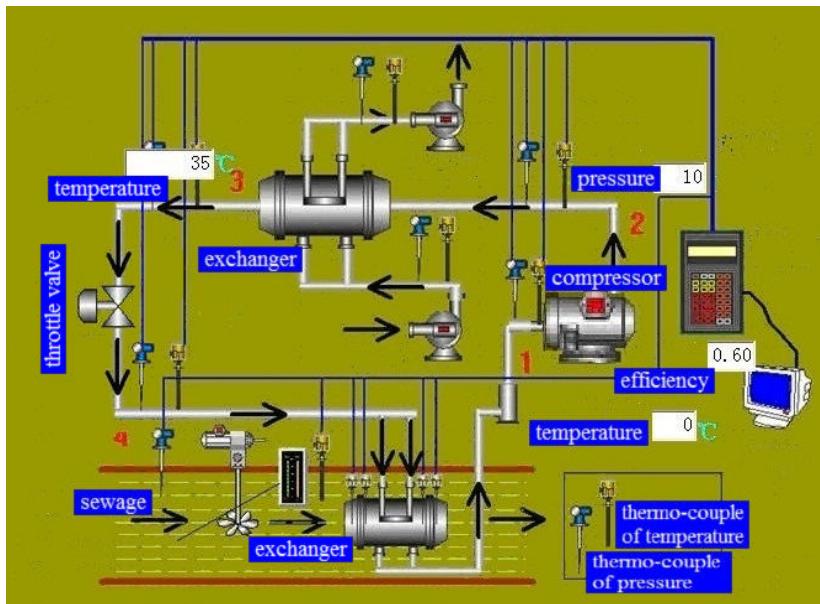


Fig. 3. Platform of CO₂ sewage source

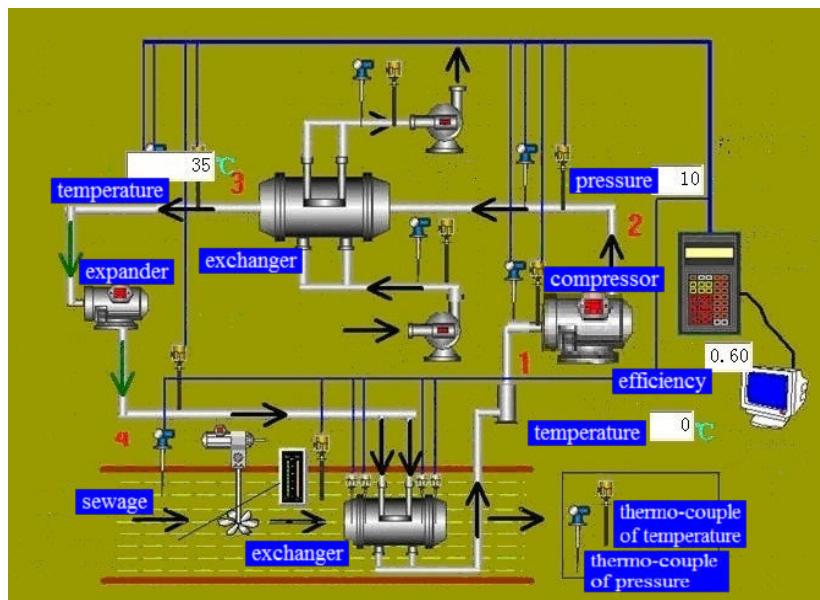


Fig. 4. Calculation platform of the two cycles

4 Results Analysis and Discussion

Fig. 5 shows the change of cycle performance with discharge pressure. In transcritical CO₂ cycle with a throttle valve, there is an optimum discharge pressure corresponding to the maximum cycle performance during in range of pressure. The reason is the gas cooler outlet temperature, evaporating temperature and compressor performance and other factors have a great impact on the optimal pressure. Under the same conditions, the cycle performance with an expander is better than the throttle cycle. With increasing of the compressor efficiency or the expander efficiency, regardless of the throttle cycle or the expander cycle, the cycle performance both of them increases.

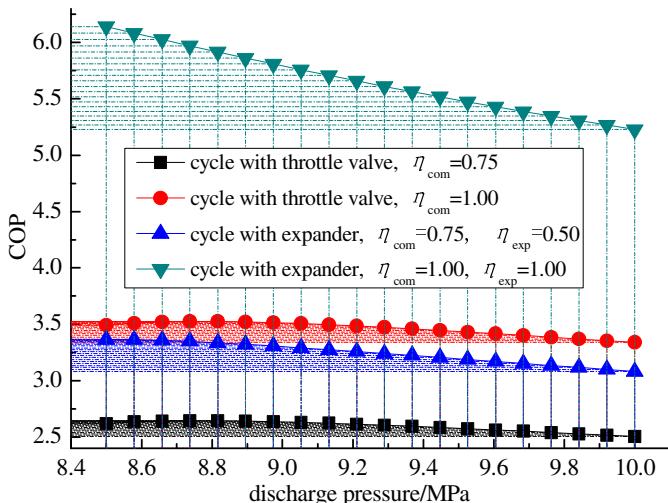


Fig. 5. COP with discharge pressure

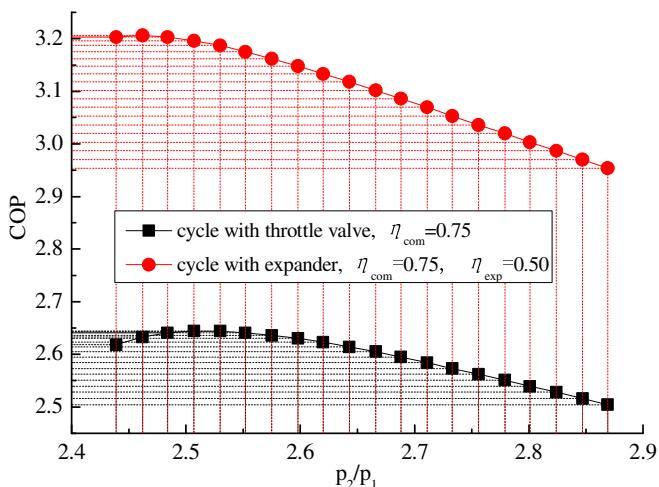


Fig. 6. COP with pressure ratio

When the expander efficiency is 0.5 and the compressor efficiency is 0.75, the COP with an expander cycle is lower than the throttle valve cycle under the conditions of entropy compression. The results show that the efficiency of compressor and expander has an effect on system performance. At the same time, the results also show that transcritical cycle with an expander has an expander efficiency limit values. Under the limit value, the expander cycle performance is not better than the throttle cycle. Only in the expander efficiency value greater than the limit value, the expander system has a significant superiority.

Fig.6 shows the change of cycle performance with compressor pressure ratio. From Fig.6, with increasing of the pressure ratio, the COP of two kinds cycles increase gradually and there is a maximum performance corresponding to the optimum pressure ratio. Under same contrast conditions, the performance of the expander cycle is better than the throttle valve cycle. In throttle valve cycle, the optimal pressure ratio is 2.53 and the maximum system COP is 2.64. In expander cycle, the optimal pressure ratio is 2.46 and the maximum system COP is 3.21.

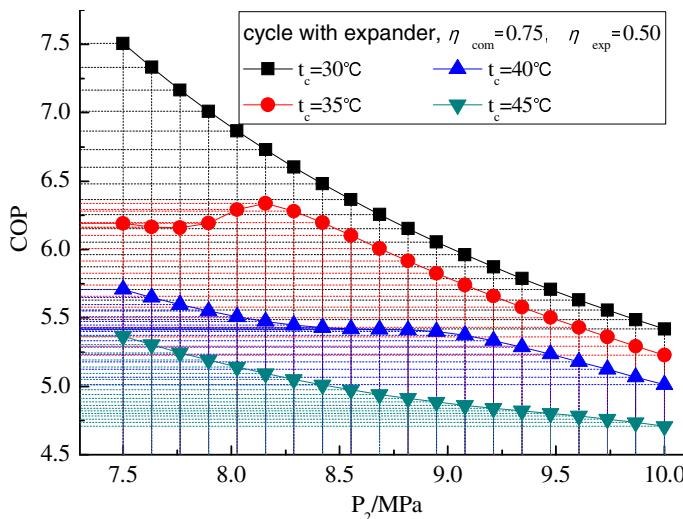


Fig. 7. Change of COP with pressure

In the ideal expander cycle (the efficiency of the compressor is 100%, the efficiency of expander is 100%), with increasing of the gas cooler outlet temperature, the CO₂ expander cycle will not present the optimum high pressure, which can be obtained from Fig. 7. As gas cooler outlet temperature is 35°C, the COP increase with the high pressure, then increase, and in a pressure to achieve maximum. As gas cooler outlet temperature are 30°C, 40°C and 45°C, with the high pressure increasing, the performance of the cycle COP declining, which does not exist the optimal high pressure.

Fig. 8 shows the changes of COP with evaporation temperature. From Fig. 8, the COP approximate linear relationship with the change of evaporation temperature, and the COP increase with increasing of evaporation temperature. Therefore, the maximum heat corresponds to the maximum evaporation temperature, but also leads to the decline of the exhaust pressure.

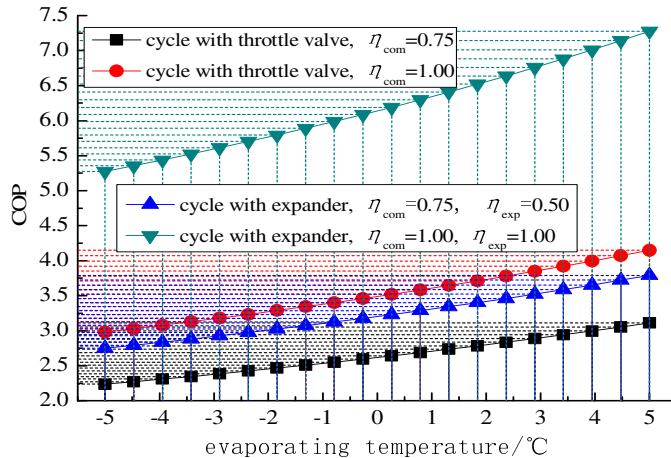


Fig. 8. Change of COP with temperature

Fig. 9 shows the outlet temperature in gas cooler effects on performance COP. Fig. 9 shows the cycle performance decline gradually with increasing of gas cooler outlet temperature. In the range of gas cooler outlet temperature, the cycle performance with an expander is better than the cycle with a throttle valve.

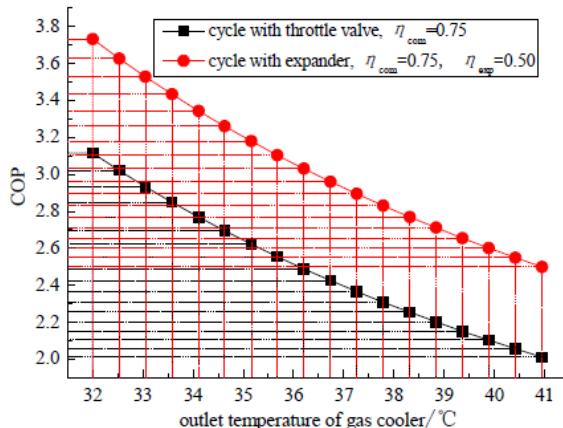


Fig. 9. COP with outlet temperature

5 Conclusions

Based on the problems of freon refrigerants ODP and GWP, the mathematical model of CO₂ transcritical sewage source heat pump cycle with a throttle valve and an expander were established. Based on Visual Basic program software, the performance analysis platforms were also developed.

(1) Regardless of the transcritical cycle with a throttle valve or an expander cycle, there is an optimum discharge pressure. The cycle with an expander has an efficiency of expander limitation value, which is the determination value of the throttle valve cycle and the expander cycle.

(2) With increasing of evaporating temperature, the performance of the two kinds cycles increase. With decreasing of outlet temperature of gas cooler, the performance of the two kinds cycles decrease.

(3) Under the same comparison conditions, the average cycle performance with an expander is better than with a throttle valve cycle.

Acknowledgements. The authors acknowledge the support by the Higher Education Major Project of Education Commission of Hebei Province under Grant ZD2010105.

References

1. Calm, J.M.: Resource ozone and global warming implications of refrigerant selection for large chillers. In: Proceedings of the 22nd International Congress of Refrigeration, Beijing, pp. 1–9 (2007)
2. Tamura, T., Yakumaru, Y., Nishiwaki, F.: Experimental study on automotive cooling and heating air conditioning system using CO₂ as a refrigerant. *Int. J. Refrig.* 28(8), 1302–1307 (2005)
3. Brown, J.S., Yana-Motta, S.F., Domanski, P.A.: Comparative analysis of an automotive air conditioning systems operating with CO₂ and R134a. *Int. J. Refrig.* 25(1), 19–32 (2002)
4. Zapusek, U., Lestan, D.: Grant.: Functioning and toxicity of artificial soil mixtures with metal-bearing sewage sludge. *Ecol. Eng.* 37(12), 1977–1982 (2011)
5. Agrawal, N., Bhattacharyya, S.: Experimental investigations on adiabatic capillary tube in a transcritical CO₂ heat pump system for simultaneous water cooling and heating. *Int. J. Refrig.* 34(2), 476–483 (2011)
6. Stijepovic, M.Z., Linke, P.: Optimal waste heat recovery and reuse in industrial zones. *Energy* 36(7), 4019–4031 (2011)
7. Sarkar, J.: Cycle parameter optimization of vortex tube expansion transcritical CO₂ system. *Int. J. Therm. Sci.* 48(9), 1823–1828 (2009)
8. Wang, H., Ma, Y., Tian, J., Li, M.: Theoretical analysis and experimental research on transcritical CO₂ two stage compression cycle with two gas coolers (TSCC + TG) and the cycle with intercooler (TSCC + IC). *Energy. Convers. Manage.* 52(8–9), 2819–2828 (2011)
9. Yang, J.L., Ma, Y.T., Liu, S.C.: Performance investigation of transcritical carbon dioxide two-stage compression cycle with expander. *Energy* 32(3), 237–245 (2007)
10. Yang, J.L., Ma, Y.T., Li, M.X., Guan, H.Q.: Exergy analysis of transcritical carbon dioxide refrigeration cycle with an expander. *Energy* 30(7), 1162–1175 (2005)

Information Retrieval Services Based on Lucene Architecture

Hang Li, Wanlong Li, Guochun Wang, and Xinyi Peng

College of Computer Science and Engineering Changchun University of Technology
Yan'an Road 2055 Changchun City Jilin Province
lihangccut@sina.com

Abstract. Lucene full-text retrieval technology is widely used in the field of information retrieval, it is an excellent, open source full-text indexing engine tool kit written in Java. This paper first briefly describes the inverted index mechanism of Lucene, and then analyses Lucene architecture and its index file structure, as the basis for introduction the two modules of the Lucene in detail. Finally point out the shortcomings of the segmentation module and the place that can be improved. Through its API, we can embed it into a variety of applications to develop our own search engine, or ccustom a personalized information retrieval service.

Keywords: Lucene, full-text search engine, architecture, index, retrieval.

1 Foreword

With the rapid development of Internet, Chinese electronic information that uses the Internet as the carrier grows quickly, and the information is generally presented in the form of HTML or plain text file, which is irregular, the text is generally long and varied content, it is difficult to use format specification. However, the requirements to the query are increasingly high. The computer indexing retrieval and Boolean logic function to retrieve can no longer satisfy people's needs.

Full-text retrieval system becomes more and more welcome because of its powerful functions and easy operation .Full-text database software is specifically used to store and query these electronic documents, and supports query online. The core of full-text database is full-text retrieval, and full text retrieval can composite manages large text, voice, images, moving images and other unstructured data.

Lucene is a very popular, open source and full-text search engine tool kit, not a full text search engine. It only provides a framework. Users can accord to the actual personal need of their application to build a complete of full-text search engine.

2 Lucene Instruction

Lucene is a subproject of the Jakarta project group apache software foundation and an open source text search engine tool kit. That is to say, it is not a complete full-text

search engine, but full-text search engine architecture. Lucene provides a complete query engine, indexing engine and part of the text analysis engine (two Western languages: English and German), and it can be easily embedded into the various applications to implement the full-text index and retrieve. The interface design of Lucene API is more general, and its input and output structure looks like the tables, records and fields in database, so many traditional application files and database can be quite easy mapped to storage structure and interface of the Lucene. The codes of Lucene are fully implemented in Java, not configured. It only supports plain text file index and search, not responsible for the else files extracted from other formats files.

The core technology of the full-text search is recording information of all the basic elements appeared in the source document and storing this information into the index repository. Lucene adopts the data structure of inverted index [7], storing the mapping of the storage location where a keyword is stored in a document or a set of document, and the document is all documents that contain the keyword and contains some auxiliary information such as the times of appearance and the location information of the keyword appeared in the document. The information can also be used to sort the search results, and the storage location can be sorted in accordance with increasing order of the document number, and finally compressed and saved as an entire data. The inverted index table of Lucene is shown in Table 1.

Table 1. Inverted Index Table

keyword: T ₁	{D ₁₁ ,TF ₁ ,<loc ₁ , loc ₂ ...>} {D ₁₂ ,.....}
keyword: T ₂	{D ₂₁ ,TF ₂ ,<loc ₁ , loc ₂ ...>}.....
.....
keyword: T _n	{D _{2n} ,TF _n ,<loc ₁ , loc ₂ ...>}.....

3 Lucene Architecture and the Main Modules

3.1 Lucene Architecture

The system structure of Lucene, first-class full-text search engine, embodies a large number of object-oriented design ideas. The first is to define a index file format that is independent of the platform, followed by the manipulation that design the system's core components as abstract classes and the concrete platform part as abstract classes too; in addition, the parts interrelated to the specific platform such as file storage package are encapsulated as classes. After a layer of processes of object-oriented programs, and ultimately produce a low coupling, high efficiency, easy to secondary development search engine system [2].

All source codes of Lucene system are divided into seven modules, expressed through package in Java. There are five main modules: org.apache.lucene.analysis, org.apache.lucene.index, org.apache.lucene.search, org.apache.lucene.document, org.apache.lucene.queryParser. Each module has a specific function. Theirs function are as follows:

(1) org.apache.lucene.analysis: Lucene language parser. It includes a complete set of analysis systems as well as an extended structure, mainly used for word segmentation. Word segmentation is mainly realized by the extended classes of the analyzer, Lucene comes with a class StandardAnalyzer, and we can refer to the realization of the class to establish our own word parser class, such as the Chinese parser.

(2) org.apache.lucene.index: Index management is its job. Similar to the table structure of the relational database, the main function is to supply the warehouse with the read and write interfaces. We can find a warehouse, add, delete and read the records through the package.

(3) org.apache.lucene.search: This module mainly supply search interface. Through this package, we can enter a query and then get query result sets. If we define the query rules in coordination with the package org.apache.lucene.queryParser, the system will support the complex query such as and, or, not.

(4) org.apache.lucene.document: This package is to manage the document structure when the index is stored, similar to a table structure in a relational database.

(5) org.apache.lucene.queryParser: Query parser. Realize the arithmetic in the keyword for query, such as and, or, not.

3.2 The File Structure of Lucene Index

The index of Lucene is composed of one or more segment that can be retrieved individually as module; each segment consists of independent several documents that may be an HTML page, an XML document or a word document; each document is composed of a certain number fields; each domain is composed of a few of term. Term is the smallest unit of the index and on behalf of a string, its position, frequency and number of times and else information in a file. Domain, an associated tuple, consists of a domain name and the domain value [1, 3]. The document is the results of extracting all the information from a file and the segment is known as the sub-index, a collection of sub-index is the index. The Table 1 shows the composition of the structure of the index files.

3.3 Two Main Modules of Lucene

Lucene mainly includes two main modules: indexing and retrieval, shown in Figure 1. The two modules of Lucene realize two main functions: first, segment the text content, build index and then storage; return results according to the query.

Indexed documents are represented by the Document object. Class IndexWriter calls the method addDocument to add the document to the index, achieves the process of creating index. Lucene's index is reversed index. When there is a request, class IndexSearcher calls the function search to search the index of Lucene. IndexSearcher also calculates the term weight and score and returns the results after sorted to the user. The documents sets returned to the user is represented TopDocsCollector.

Indexing Module. First, establish the document to be indexed. A document has a variety of information such as title, author, modification time, content. Different type

information is expressed by different Field, and then joins the Field to the document. Lucene set up index for text files, but it provides several parsers to parse a variety of different types of documents and extract text information [4].Figure 2 shows the transformation process. Take HTML files as an example, HTML parser takes some preprocessing such as filtering HTML tags in the document. Finally, output the text content.

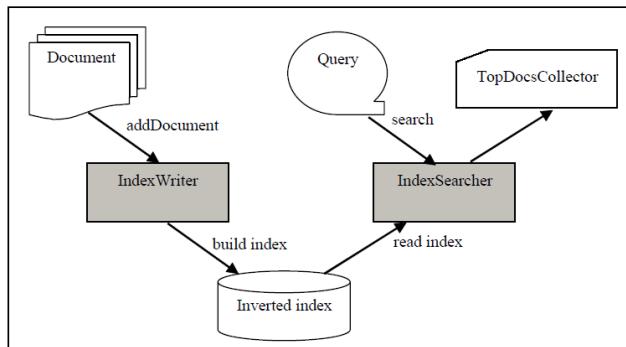


Fig. 1. Full-text search module based on Lucene

Second, use the parser of Lucene to filter the information gained by crawler and then segment. Filters act on the results formed after segmentation and get rid of some sensitive words, stop words (a, an, the, of) etc. Finally the terms are formatted. Term arises in the structure of Token. Token is a structure recovered to the indexer (lucene.index) after that language analyzer analyses the source files. Token records the information required by each index field.

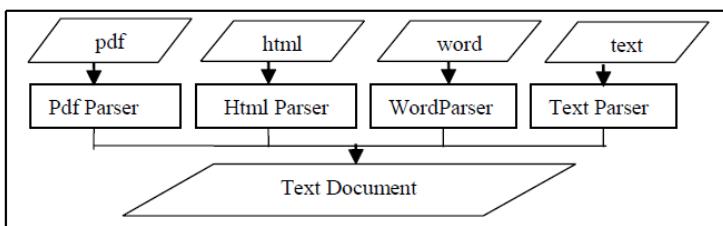


Fig. 2. The process of Document Parsing

Finally, create index after the terms are gained from segmentation. The content of the index not only contains the term itself, but the related information about the term. The codes of the index module are as follows:

```

createIndex(String fPath, String iPath, Analyzer a)
{
    File file=new File(fPath);File[]
    files=file.listFiles();
    for(int i=0;i<files.length;i++) {
  
```

```

Document doc1=FileDocumentUtils.fileDocument(files[i].getAbsolutePath());
IndexWriter indexWriter1 = new IndexWriter(iPath+"index_"+i, a, true,
MaxFieldLength.LIMITED);
indexWriter1.addDocument(doc1);
indexWriter1.close(true);}}
```

Document fileDocument(String path) {

```

File file = new File(path);
Scanner scanner=new Scanner(file);
String address=scanner.nextLine();
Document doc2 = new Document();
doc2.add(new Field("name", file.getName(), Store.YES,
Index.ANALYZED));
doc2.add(new Field("content", readFileContent(file),
Store.YES, Index.ANALYZED));
doc2.add(new Field("size",
NumberTools.longToString(file.length()), Store.YES,
Index.NOT_ANALYZED));
doc2.add(new Field("path", file.getAbsolutePath(),
Store.YES, Index.NOT_ANALYZED));
doc2.add(new Field("address", address, Store.YES,
Index.NOT_ANALYZED));
return doc2;
}

```

Lucene is a universal search engine frame. If we make crawler choose dedicated industries and fields to obtain the specialized field data when the system is in motion, formatted store and manage them, and then make use of the document module to establish corresponding document structure. Finally, we can use the full-text indexing module and search module of Lucene to make it vertical search engine, supplying personalized information services.

Retrieval Module. Search, the most important function of a search engine, directly affects the user experience on the use of search engines. The query logic of Lucene is divided into four sections:

- (1) The user enters a query.
- (2) Lucene will deliver the query to the searcher (lucene.search).The searcher calls the query parser (lucene.queryParser) included in the searcher to parse the Lucene

query parser, and then calls the language analyzer (lucene.analysis) to segment words and filter them.

(3) Searcher fast traverses the index tree, a query warehouse that has been established in advance to get the query results, does relevance evaluation to get a sorted document lists, and then displays to the users after reorganized them to the search result.

(4) Display the returned result set in the query results page. When the user clicks a piece of content, it can be connected to the original page or open the page stored in the full-text database. The codes of the search part are as follows:

```

search(String queryStr, String iPath, Analyzer a)
{String[] fields = { "name", "content" };

QueryParser queryParser = new
MultiFieldQueryParser(fields, a);

Query query = queryParser.parse(queryStr);

File file=new File(iPath);int hitCount=0;

ArrayList<Document> matchedResult=new
ArrayList<Document>();

for(int i=0;i<file.listFiles().length;i++){

IndexSearcher iSearcher = new
IndexSearcher(indexPath+"index_"+i);

Filter filter = null;

TopDocs topDocs = iSearcher.search(query, filter,
10000);

hitCount+=topDocs.totalHits;

for (ScoreDoc scoreD : topDocs.scoreDs) {

int docSn = scoreD.doc;

Document doc = iSearcher.doc(docSn);

matchedResult.add(doc);

FileDocumentUtils.printDocumentInfo(doc); }
}

```

The results of Lucene default sort in accordance with the degree of correlation. When there are the same score in two documents, the default sort is in accordance with the document ID. We can use the following settings to change the order of the results:

(1) Changing the Document boost (inspire factor): changing the size of the boost will result in the change of the Document score, thus changing the order of the Document in advance or later. Boost is used when we calculate the score of the documents, and its default value is 1.0, meaning the core is unrelated to boost. Once you have changed the default boost value, will take changes from unrelated to relevant: the greater the boost value, the higher the score of the Document.

(2) Changing the boost of the Field: it is the same as changing the boost of the document.

(3) Using Sort tool to order: When Lucene is in query; we can construct a searcher (IndexSearcher) through setting an object Sort as a parameter and appoint the sort rule at the same time.

3.4 Segmentation Module of Lucene

In the process of building index and retrieving, we use term as the basic unit. Lucene language analysis package org.apache.lucene.analysis is the shared module by the index module and retrieve module. Because we must use same analyzer, using the same segmentation rules to obtain the same results after dividing phrases or sentences, to ensure the retrieval process is carried out smoothly.

Before the document storage, and after user inputs query, we must use Lucene language parser to segment and filter them. Analyzer is an abstract class to achieve the segmentation of the text content. The function of Analyzer is mainly realized through two classes: TokenStream and TokenFilter. Tokenizer processes a single character, reads the data in the Reader object and transform them into a lexical unit. TokenFilter has the function of text filter, but we need to pay attention to the order of the use of different filters when we use them. Lucene itself comes with variety of analyzer and filters. For example: StandardAnalyzer and StopAnalyzer. The latter can filter a particular string and vocabulary and complete the uppercase turn lowercase function.

The analysis rule of Lucene to English is simple. English takes space as a sign of the word segmentation. There is no space between the words of Chinese, and in the process of segmentation, we must encounter the ambiguity problems and unknown word identification, so the word segmentation for Chinese is more difficult. Lucene provides two Chinese analyzers: CJKAnalyzer and StandardAnalyzer, respectively supplying with single word or double words segmentation mechanisms. CJKAnalyzer calls CJKTokenizer for the Chinese word segmentation; at the same time uses StopFilter to complete the filtering function, achieving the multiple segmentations of the Chinese and stop word filtering. StandardAnalyzer accords spaces, and symbols to complete segmentation and can analysis and process the numbers, letters, e-mail address, IP address, as well as Chinese characters. It also support the filter vocabulary instead of StopAnalyzer to realize filters.

Two methods can not better support the Lucene Chinese analysis processing not only at quality but efficiency. Major popular analyzers: IKAnalyzer, ICTCLAS4J, Imdict-Chinese-Analyzer, Paoding Analysis. They all inherit from the Analyzer of Lucene, processing Chinese text. Therefore, we can through the Lucene API, extend the class of Analyzer to achieve a suitable, efficient tokenize, such as segmentation based on statistic, dictionary-based segmentation and segmentation based on semantic understanding. We also can introduce some models in the process of segmentation, such as hidden Markov model, maximum entropy model.

4 Conclusions

Lucene is an open source and based on java full text information retrieval tool kit, providing an excellent opportunity for us to learn search engine. We can extend the functionality of the Lucene through its application programming interfaces and embed it into a wide variety of application programming to provide the index and retrieve engine for them or ccustom a personalized information retrieval service.

References

1. Zhou, D., Xie, K.: Lucene search engine. Computer Engineering (2007) (in Chinese)
2. Guan, J., Gan, J.: Applied research based on the Lucene text search engine. Computer Engineering and Design (2007) (in Chinese)
3. Zhao, T., Meng, X.: Chinese full-text database design and implementation based on Lucene API. Computer Engineering and Applications (2003)
4. Li, Y., Ding, H.: The research and application of full-text retrieve of Lucene. Computer Technology and Development (2010)
5. Tang, H., He, Y., Xu, X.: Distributed parallel index based on Lucene. Computer Technology and Development (2011) (in Chinese)
6. Ding, Z., Kim, M.: Research and Implementation of personalized search engine based on Lucen. Computer Technology and Development (2011) (in Chinese)
7. Zheng, R., Lin, S.: Research on Chinese inverted index technology based on Lucene. Computer Technology and Development (2010)
8. Son, J., Zhu, Y., Liu, R.: Improved full-text search tool kit based on the Lucene. Computer Engineering and Applications (2008)
9. Li, S., Ling, F., Lv, X., Shi, S.: Study on Efficiency of Full-Text Retrieval Based on Lucene. In: International Conference on Information Engineering and Computer Science (2009)
10. Hatcher, E., Gospodnetic, O.: Lucene in Action. Manning Press, Greenwich (2004)
11. Qiu, Z., Fu, T.: Develop its own search engine Lucene 2.0 + Heritrix. People Post Press, Beijing (2010) (in Chinese)
12. Bruce Croft, W., Metaler, D., Strohman, T.: Search Engines Information Retrieval in Practice. China Machine Press (2010)

Analysis on the Stability of Cavity Based on Cavity Monitoring System

Zhang Haibo^{1,2}, Song Weidong¹, and Liu Fangfang³

¹ Civil and Environmental Engineering Institute, Beijing University of Science and Technology,
Beijing 10083, China

² College of Mining Engineering, Hebei United University Tangshan 063009, China
³ Hebei Vocational College of Geology, Shijiazhuang 050081, China
zhbts@126.com

Abstract. The key of the stability numerical simulation in cavity is to grasp its 3D configuration and spatial location accurately. Based on the actual situation of Shi Rengou iron mine and cavity 3D laser detection system, this paper develops the cavity 3D model and geological 3D model by using the 3Dmine software. Numerical computation model is developed by the coupling of 3Dmine and Flac3D with the data conversion of coupled mode. Numerical analog computation is conducted, and the surrounding rock stability in cavity is analyzed according to the simulation results. The research shows that the combination of the advantages of CMS, 3Dmine and Flac3D provides a new and effective way for numerical simulation research on the stability of rock engineering.

Keywords: Cavity, Modeling, Surrounding Rock Stability, Numerical Simulation.

1 Introduction

With the rapid development of computers, numerical simulation technology has made good progress and has been widely applied in the field of mining and geotechnical engineering research [1-5]. However, the numerical simulation research on the stability of a single cavity is far more than the overall numerical simulation research on cavity group. In the meanwhile, the traditional numerical simulation analysis was conducted by the determined cavity scope in the process of design, and the boundary of cavity is regarded as regular. However, the actual cavity is in an irregular shape due to the influences of measuring, vibration of blasting and falling of surrounding rock, which is far different than the designed one. Taken the above factors into account, the precision and reliability of stability of cavity numerical simulation result will be seriously affected in the process of numerical simulation.

2 CMS Detection Technology

The basic components of CMS include laser scanning head, control box, handheld controller, support bar frame and data-processing software.

CMS detection is mainly conducted through the 360° rotating scanning head for implementation of integrated laser rangefinder. At the same time, while scanning head is 360 ° rotating, it collects distance and angle data. Scanning head will be repeated automatically through raising its elevation according to setup angle by operation personnel after completed 360° rotating, and then collects more data on the points of rotation circle, until completes all detection work [6].

While scanning the empty area by CMS, the distance that from the point to laser scanned from a certain angle in advance will be saved to TXT file automatically. Then after exporting this file, it will be processed by CMS PosProcess Software that contained in CMS software. Input anchor point measured by total station, then export DXF files or XYZ coordinates file directly. These files can develop cavity 3D physical model through some mine modeling software.

3 Build Up the Cavity 3D Model

By using 3Dmine software, dxf or xyz file that got by CMS data preprocessing can be imported to 3Dmine software directly. After entity edit and validation, the final physical model can be built up.

3Dmine and FLAC3D data coupling mode is converted the model data built up by 3Dmine model into the acceptable FLAC3D data format.

The establishment of the earth's surface and whole 3D model need to import the topographic and geologic map of mining area to 3Dmine, then generate the surface physical model by 3Dmine. By using entity surface grid in the physical tools, elevation grid points on surface can be got. Select the grid points, copy it into Excel, 3D grid coordinates of the surface can be got. After further operation, required Flac3D command stream file can be generated, and then the computational model can be developed.

The establishment of numerical computation model of ore body and cavity need to build up the block model on the basis of 3Dmine at first. The centroid point file exported from block model includes center coordinates, edge length, lithology of each cell block. The centroid point file data is converted into command corresponding data in stream file of Flac3D, this process is as follows [7-11]:

(1)The central coordinates of each cell block in the output file plus or minus half length of each side of cell block can get basic control point P1, the P2 and P3 points X, Y, and Z coordinate of the hexahedral.

(2)By using three control points and group keyword, establish grouping. First establish the grouping of ore bodies, and then establish the grouping of cavity.

(3)Organization and formation of the command stream file that can be read by Flac3D.

On the basis of analysis of the numerical simulation model on the overall region, read the grouping command stream file of the ore body and cavity, the final numerical simulation analysis model can be got.

4 Computation Model and the Selection of Physical Mechanics Parameter

Due to the mining cavity is not disposed during the exploitation of Shi Rengou iron ore in recent years, the large area of mining cavity has formed at 0M to -60M level. the existed cavity selected by computation model includes: the csj2#~12#, xj1#~4#, 39#,40#, and illegal cavity fck1#~fck4#, amount to 21. The table1 shows the range of numerical simulation. Be confined to the limited space, this article gives some simulation results.

Table 1. The range of numerical simulation

Xmin	Xmax	Ymin	Ymax	Zmin	Zmax
3150	3510	7000	7500	-100	181.75

X and Y direction size are 4m in model, because the cluster analysis in cavity is concentrated at -60m~10m level, the density of grid within this range in Z direction is increased, other areas is sparse properly. The physical and mechanical parameters are used in model in Table 2.

Table 2. Experimental data table of physical and mechanical parameters

Name	Bulk density (g/cm ³)	Compressive strength (MPa)	Tensile strength (MPa)	Internal			Elastic modulus (MPa)	Poisson's ratio
				Cohesion (C/MPa)	friction Corner (φ)			
M1 ore body	3.58	99.44	4.64	1.72	41.11		8.03	0.21
M2 ore body	3.46	130.77	5.64	1.59	45.36		7.59	0.20
Surrounding rock	2.74	141.58	6.18	2.38	46.82		6.98	0.26

5 Formed Initial Stress Field

The cover depth of simulated cavity is quite shallow, so select elastic solving process to generate the initial stress field. The bottom of the model should be fully constrained, and the four sides should be fixed. The earth's surface is regarded as the free surface, and the self-weight stress of the rock mass is taken into account only. Constitutive model is set up as the elastic model, solving, then replaced by the elastic-plastic model, continue to solve, get the initial stress field setting final.

The excavation of this cavity numerical simulation is conducted by actual mining sequence, but the time of each layered mining is not continuous. In order to ensure the accuracy of the calculation results and save workload, in the process of simulation, assume each room is excavated at one time. In this process, the detected cavity model is built up in accordance with the actual shape. As for the undetected cavity, it should be excavated in accordance with design size. The illegal cavity model is set up in accordance with geophysical prospecting data. At the beginning, excavate the illegal cavity then excavate -60m level of the middle piece of cavity in sequence.

6 Analysis of Simulation Results

The simulation results are mainly used to analyze the distribution of displacement field, stress field and the plastic zone after excavation. After calculating the model, slice it and make sure the profile across the central position of cavity. Analyze the closed cavities at the same profile map, and make sure each cavity is sliced.

(1) The displacement field analysis.

Fig.1 and fig.2 are Z, Y displacement map of csj-2#, 11#, 12#and 8# cavity. csj-8 # and csj-12 # cavity have been completed. The maximum Z displacement is -6.45cm on roof. The maximum Z displacement is -6.4cm on csj-2 # roof. The jambs between csj-12# and csj-11# cavity gradually shift from top to bottom Y displacement, which point to 11 # cavity at the top and point to 12 # cavity at the bottom. The maximum displacement is 1.28cm, which points to 12 # cavity. The jambs displacement between 11#和2# cavity is 0.75mm, smaller, which points to 2# cavity.

(2) Stress field analysis.

The impact on surrounding rock made by excavate cavity is non-linear and irreversible, which makes the stress of surrounding rock in the initial stress state redistribute, finally reaches new equilibrium.

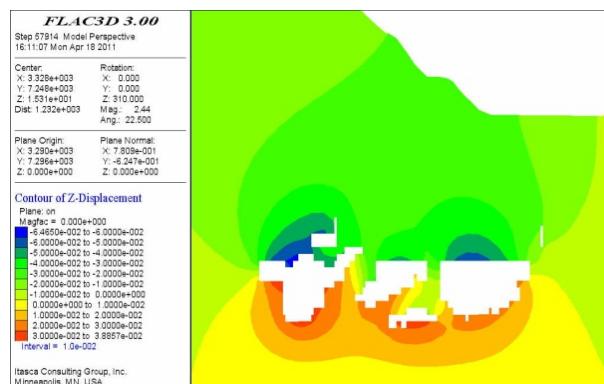


Fig. 1. The displacement of Z in group of cavity

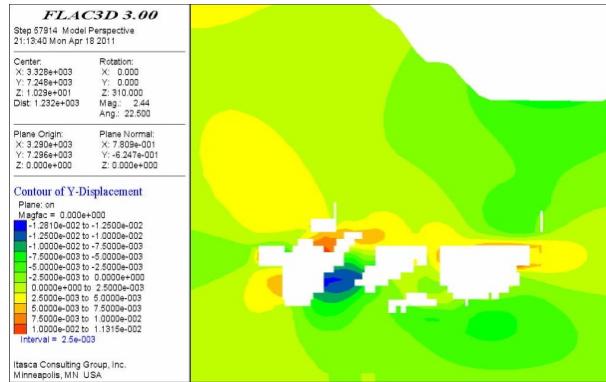
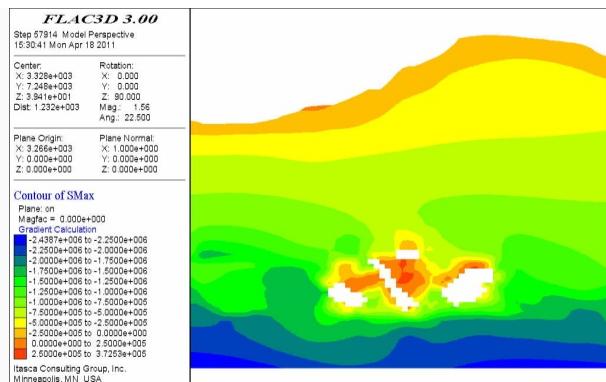
**Fig. 2.** The displacement of Y in group of cavity

Fig.3 and fig.4 are minimum and maximum principal stress diagrams of 8 # cavity, the middle cavity is 8 # cavity, the two sides are 11 # and 12 # cavity. It can be seen from the figures that tensile stress zone appeared with large scale in the upper and lower positions of the cavity, the size is 0.37Mpa, which appeared in the 8 # cavity and the roof position in 11# and 12# cavity. The minimum principal stress is 0.5Mpa appeared on 8 # cavity roof, the maximum principal stress is 9.0Mpa in 8 # cavity, which appeared at the bottom plate of the 8 # cavity. The maximum principal stress is 9.18Mpa on the left wall of the 12 # cavity.

(3) Analysis of the plastic zone.

Fig.5 is plastic zone map for the csj-8 #, 11 # and 12 # cavity, the expansion of plastic zone in this region is very serious. The surrounding rock between these three cavities all appeared the plastic zone, and the majority is on plastic stage. Fig.6 is plastic zone for csj-12 # cavity and illegal cavity. The 12 # cavity wall rock appeared plastic zone as well as jams on illegal cavity, which are in danger. The two illegal cavities all appeared plastic zone, which are dangerous too.

**Fig. 3.** The minimum principal stress in 8# cavity

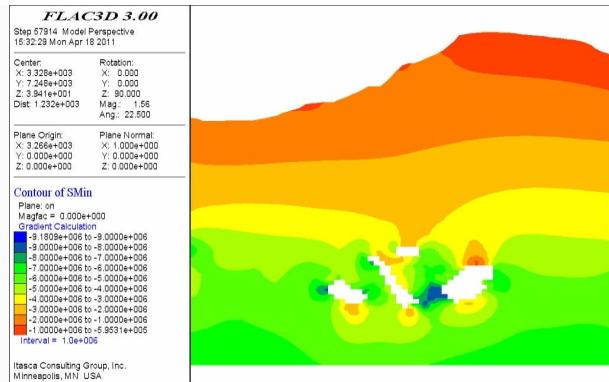


Fig. 4. The maximum principal stress in 8# cavity

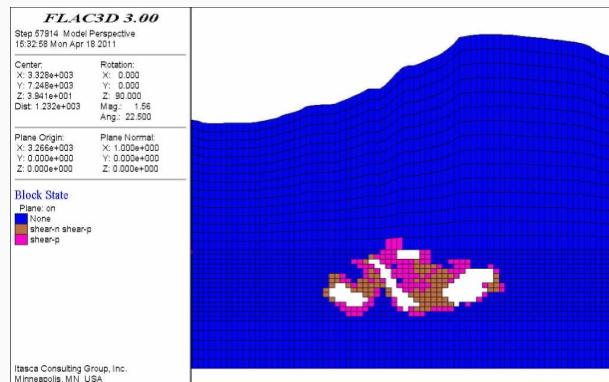


Fig. 5. The plastic zone in 8 # and other areas

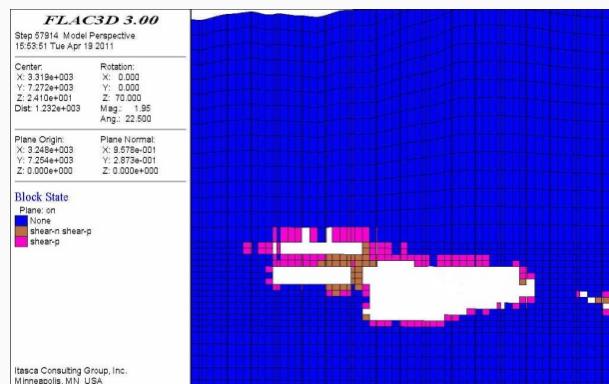


Fig. 6. The plastic zone in 12 # cavity

7 Conclusions and Suggestions

The utilization of CMS equipment to detect the cavity and the establishment of 3D model by 3DMine mining software can provide an accurate data base for the stability of the cavity numerical simulation. Achieve 3Dmine and Flac3D numerical simulation coupling, overcome the modeling difficulties effectively on the traditional numerical simulation analysis, and open up a new way for the numerical simulation of stability of geotechnical engineering.

After the cavity excavation, the displacement of the cavity mainly face to the roof of the Z displacement and the lateral displacement of pillar is smaller; the minimum principal stress of the cavity roof is very small, which explains the cavity roof close to tensile stress region gradually, some cavity even appear tensile stress. From the analysis of distribution of plastic zone, the cavity in this area appear a "group effect" that makes cavity influence each other and aggravates the damage of surrounding rock, which has a greater impact on the stability of the cavity. Therefore, this paper suggests dealing with the cavity in time.

References

- Lin, H., Cao, P., Li, J.T.: Automatic Generation of FLAC Model Based on SURPAC. *J. of China University of Mining & Technology* 37(3), 339–342 (2008)
- Kou, X.Y., Jia, M.T., Wang, L.G.: Evaluation and analysis of stability of mined-out area based on the CMS and DIMINE-FLAC3Dcoupling technique. *J. of Mineral Engineering Research* 25(1), 31–35 (2010)
- Luo, Z.Q., Wu, Y.B., Liu, X.M.: FLAC3D modeling for complex geologic body based on SURPAC. *J. of Rock and Soil Mechanics* 29(5), 1334–1338 (2008)
- Zhang, Y.H., Wang, Z.Q., Yang, Y.H.: Flac modeling and simulation for underground mining on SURPAC. *J. of Underground Space and Engineering* 7(2), 306–310 (2011)
- Yang, H.J., Dong, C.J.: Numerical Simulation of Roadway Stability in Deep Mining based on ANSYS. *J. of Coal Technology* 28(4), 50–51 (2009)
- Cavity Monitoring System Operation Manual. Optech System Corporation, Canada (2008)
- Zhang, Y.P., Cao, P.: Numerical Simulation on Stability of Complicated Goaf. *J. of Mining & Safety Engineering* 27(2), 233–238 (2010)
- Li, M., Zheng, H.C.: FLAC3D Analysis of Goaf Stability of Jia-zhuang Gypsum Mine. *J. of Non-Metallic Mines* 32(2), 54–57 (2009)
- He, Z.M., Peng, Z.B.: Numerical analysis for roof stability of double gob area after excavation by FLAC3D. *J. of Central South University (Science and Technology)* 40(4), 1065–1070 (2009)
- Sun, G.Q., Li, J., Hu, X.B.: FLAC3D Based Stability Analysis of Mined out Area. *J. of Metal Mine* 368(2), 29–32 (2007)
- Tian, X.G., Chen, H.M., Chen, S.S.: Three-dimension Finite Element Analysis on Mined - out Area Stability. *J. of Mining R & D* 27, 32–34 (2007)

LED Package Based on MATLAB Program

Zhiqin Huang¹, Xiuyun Zheng², and Fengling Liu¹

¹ School of Mathematical Sciences, University of Jinan, Jinan, 250022, China

² School of Science, Xi'an University of Architecture and Technology Xi'an, 710055, China

hzq-1230@163.com

Abstract. The light distribution simulation of LED is very important before the production. it is necessary to do meaningful light distribution simulation before the production of LED. At present, in the market, there are very few companies which can produce LED chips, because of less investment, low threshold into line, most of the companies are doing package and application of LED, it is needed to do the design of LED light distribution. In this paper the MFC simulation program by MATLAB program can run in every operating system. Moreover, the design of the program is of high practicability. The research on problems of LED light distribution will bring meaningful and valuable reference and guidance for some enterprises.

Keywords: LED, MFC program, MATLAB program.

1 Introduction

In order to improve the light extracting rate of LED, most of the LED products' packages set an optical lens between chip and air by using epoxy resin currently, luminous efficiency can be improved by about 50%, but the package still uses the concept of electronic packaging, only placing a component in the so-called "package". Optoelectronic packaging should consider its requirements for optical properties. Now the questions are that the same model LED optical axes have not been positioned, and the different types LED lens systems lack the original marking and positioning for the different demands and shapes, so that measurement error of 10% or more, such a big error in optical measurement can be hardly worthy of belief [1]. In this paper, MFC procedures coded by TRACEPRO software can be very easily intuitively showing optical properties of LED model, and can be used as the reference basis of designing practical LED products. After the program run, two text files are created characterizing LED emitting property. These files have important data and importing the data to MATLAB [2] can visually show optical properties of LED.

2 Led Package's Optical Computer Simulation, Results Analysis and New Design

This section uses MFC procedures and TracePro [3] simulate several common LED encapsulation structures, and change some parameters of sealing structure to draw

affecting LED light intensity distribution curve, the emitting light distribution, the angle of the optical properties of the factors. The simulation results express in three forms: light distribution chart, photometric distribution curves, and light intensity distribution of three-dimensional figure. Analyzing the results of the simulation, and then design LED.

2.1 The Selection of LED

In this section, common traditional packaging of LED is analyzed, the cylindrical dome type packaging of exterior is selected, the reflective bowls are divided into large trapezoidal angle large bowl, trapezoid large angle small bowl, trapezoidal small angle bowl, chip divided into point light source (MFC program) and light source (TracePro) [4]. Respectively studying the different inserting depth of the same reflective bowl, the same depth of the different light bowl LED optical properties, finding out the rule of LED features and design characteristics. In the table 1, there are LED parameters of this section studied.

Table 1. The parameters of LED

Number	LED configuration	thick(radius mm)	height of the columns(mm)	Reflective bowl appearance	inserting depth (mm)
1	dome columns	3	6	large Angle big bowl	0
2	dome columns	3	6	large Angle big bowl	2
3	dome columns	3	6	large Angle big bowl	4
4	dome columns	3	6	large Angle small bowl	0
5	dome columns	3	6	large Angle small bowl	2
7	dome columns	3	6	large Angle small bowl	4
9	dome columns	3	6	small Angle bowl	0
10	dome columns	3	6	small Angle bowl	2
11	dome columns	3	6	Small Angle bowl	4

The size of the reflective bowl is: large angle bowl's opening angle is 90 degrees, known as the bowl wide 1.6mm, high 0.45mm, for small bowls, bowl wide 1.1mm, high 0.25mm. Small angle bowl's opening angle is 70 degrees, high 0.32mm, wide bowl 1mm.

2.2 LED Simulation and Optical Properties Graphs

Fig. 1 (a) is polar coordinates distribution curves of TracePro, fig. 1 (b) is orthogonal coordinate distribution curves. The normal distance between absorption surface and LED is set 10cm.

2.2.1 The MFC Programs and Tracepro Simulation LED Chart Is Following

(1) known from Fig. 2(a), using large Angle big bowl, for 0 insertion depth of the LED package, light intensity distribution is very concentrated, light-emitting angle is

from -10 degrees to 10 degrees, bright central spot, strong light direction. Adopting the large angle small bowl and small angle bowl, so is the 0 insertion depth of LED light intensity distribution.

(2) A large Angle big bowl, 2 inserting depth

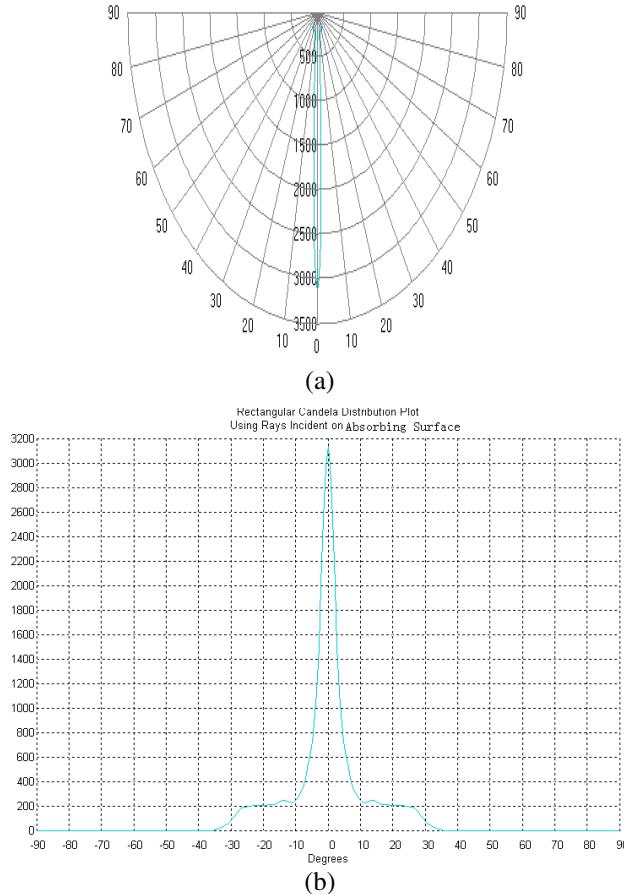


Fig. 1. Large angle big bowl, LED optical properties of 0 depth

2.2.2 The Analysis of LED's Light Intensity and Light Distribution Curve

In the simulation, the refraction rate of the medium is different; the different packaged epoxy resin shape will influence LED's light intensity and light distribution curve. In this paper, the depth of penetration of the chip and the reflection bowl size are studied.

(1) Same LED and different chip insertion depth: In section 2.2.1, we make the MFC program and TracePro software simulation for the main three LED packages included large angle bowl, large angle small bowl and small angle bowl. From mimic diagram of different chips insertion depth with the same kinds of LED, it can be

summarized as follows: When the insertion depth of chips is relatively shallow (0, 2 insertion depth), LED light-emitting angle is very narrow (-10 degrees to 10 degrees), points of light are very strong and central spot is very bright. When the chip is inserted into more depth (4, 5 insertion depth), LED light-emitting angle is wider (-40 degrees to 40 degrees), the light curve is divided into three peaks (main peak and two secondary peak). With increasing insertion depth of the LED chip peak position and the second peak position relocate, the main peak width increases, the peak decreases, and the center brightness reduced.

Based on the above analysis, it can be concluded: For the LED light intensity at the center of the stronger and more focused, a shallow insert assembly is used; to make the LED light evenly soft, divergence angle, and wide field of vision, deep into the assembly should be used.

(2) Different LED and same chip insertion depth: Trapezoidal bowl and large-angle wide-angle trapezoidal bowl [5]: taking 4 insertion depths, we can obtain: because they are the same angle reflector bowl of Chu but different high, the chip in the relative position of reflective bowl is different. Relatively speaking the mouth of bowl becomes smaller and that of small bowl bigger. Chips emit photons after reflection through the reflector, the photon emitted from the bowl Chu concentrated, and the small bowl of scattered photons.

Based on the above analysis, it can be concluded that with the use of large-angle trapezoidal bowl of LED, the light emitted form the light distribution is more concentrated, and peak intensity is stronger.

2.3 New Design of LED Package

Section 2.2 analyzes the different LED chips into the same depth, same LED chips of different insertion depth, the same insertion depth of the reflector bowl of the different optical characteristics of LED, and draws a few conclusions.

In this paper, a multi-chip LED is designed. It is available in the market that sending three primary colors of red, green and blue LED use the same LED in the multi-chip package. There is a LED with a reflector for the bottom radius of 1mm, the top surface of the radius of 2mm, high-1mm, bottom cylinder of radius 3mm, high-5mm, 3mm hemispherical radius. Three different chips into the depth of the (insertion depth1, 2, 4) are simulated, in which the number of light-emitting chip is 5, as shown in Fig. 2(a), insertion depth 1, as shown in Fig 2(b), insertion depth 2, as shown in Fig 3(a), insertion depth 4, as shown in Fig. 3(b).

By Fig. 2 and Fig. 3, it can be seen that using the light into light assembly of LED, the center is powerful, light concentration, peak range of small and strong directional; using deep into the assembly of LED, the center peak is lower than the peak on both sides, center of light intensity and light scattered light angle. Therefore, it verified the characteristics of LED of the different insertion depth of the same LED.

The advantages of multi-chip LED structure are that a large luminous flux, luminous intensity is strong, more uniform light and suitable for high power LED. Its multi-color property can be used as outdoor advertising, LED display and so on.

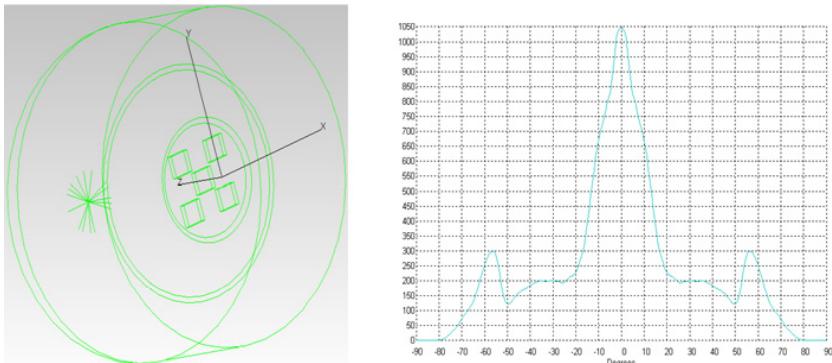


Fig. 2. (a) new designed LED (b) new designed optical features

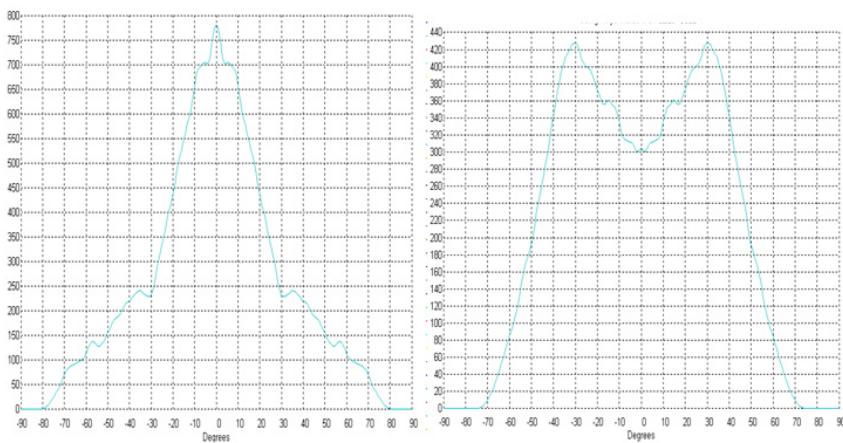


Fig. 3. New designed optical features of insertion Led (a) insertion depth 2 (b) insertion depth 4

3 Conclusions

Through comparative analysis of simulation results for the traditional package (cylindrical dome) LED different insertion depths and different reflector, the following can obtain:

- (1) To obtain small light-emitting angle, good direction, light intensity, light degree of light concentration of LED, it requires the use of light into the small bowl package.
- (2) To obtain large light angle, light scattered, soft, uniform light intensity of LED, it needs to adopt wide-angle deep into the large bowl package

The paper also simulates multi-chip LED structure. Multi-chip LED and single-chip LED has the same optical properties, but its greater luminous intensity, stronger light intensity. That is because its multi-color can be used for display.

According to the MFC that MC designed, to a certain extent, the LED's optical properties can be simulated. But sometimes the simulated LED spatial distribution curve is not symmetrical. In order to make the program more perfect, C++ language design should be used; MFC program will be further improved and optimized.

Acknowledgments. This paper was supported by a soft science project of Shandong province (No. 2012RKB01101).

References

1. Long, X., Zhou, J.: Multifunctional LED Parametric Test System. *Semiconductor Optoelectronics* 28(2), 179–182 (2007)
2. Yan, J., Yu, Y.: LED's Optical Encapsulation Structure Design Based on Monte Carlo Simulation Method. *Chinese Journal of Luminescence* 25(1), 90–93 (2004)
3. Chen, Q., Liu, H., Li, J., Shen, Y.: Study on the light source model of LED. *Lamps & Lighting* (03), 14–17 (2005)
4. Lv, Z., Zhao, Z., Fan, Q., Lv, L., Ma, C., et al.: From Distribution light curve. *China Lighting* (2004)
5. Chen, Y.: *The Manufacturing Technology and Application of LED*, 2nd edn. Electronic Industry Press, Beijing (2009)

Design and Implementation of Fuzzy Query System Based on Ajax

Fenghua Liu^{1,*}, Aiming Wang¹, Liangliang Chen², and Guanghui Wei²

¹ Department of Computer Xinjiang Polytechnical College, Urnmuqi Urumqi, 830091, China

² School of Water Resources & Civil Engineering Xinjiang Agricultural University Urumqi,
830052, China

Xj_1fh@sina.com

Abstract. In the standardized code system, the Pingyin retrieval method is used to solve the problem of newly added accurate matching code. In the Pingyin retrieval method, Ajax technology is adopted to implement the data query and data enter under the condition of unrefreshing pages. This system used dual fussy query method, and solved the problem of query efficiency and accuracy while there exists massive unchecked data. After being used, the system had shown very good performance.

Keywords: Fuzzy query, Structured Query Language, Regular Expressions.

1 Introduction

With the rise of Internet technology B/S model system has been large-scale application. There are some inadequacies in Traditional Web:

(1). Exclusive request. If a request has no completed response, the next request can not be sent. During the wait for the server response, the user's browser is blank.

(2). Page refreshed frequently. Traditional Web applications are basically corresponding request - Page mode, Each request must discard the current page to reload a new page. Frequent page refreshes will not only allow users in the waiting, but also increase the burden on the server.

Ajax [1] (Asynchronous the JavaScript and XML) was born to compensate for the deficiencies of the traditional Web. Ajax is a group of interrelated web development techniques used on the client-side to create asynchronous web applications. With Ajax, web applications can send data to, and retrieve data from, a server asynchronously in the background without interfering with the display and behavior of the existing page. Data is usually retrieved using the XMLHttpRequest object. The browser sends a request, without waiting for a response from the server, can continue the original operation. When the server's response is completed, the client using JavaScript function corresponding data is loaded into the browser. Ajax applications, especially applies to

* Corresponding author.

the Web applications such as frequent interaction and data reading, etc. There are two advantages of using Ajax technology:

- (1). without refreshing to the updated page will give users a better experience;
- (2). Transferring the work of the server to the client make the burden on the server and bandwidth to be reduce, and make the space and bandwidth rental costs to be saved.

2 Fuzzy Query Based on Ajax Technology

2.1 The Problem of Inputting Standardized Code to System

At present, most of the management information system is the B/S mode, so users can access and manipulate data in the network terminal. Standardized coding provides a convenient for the data statistics and query operations in database management. For example, in the management system of college graduates, the unified coding has been prepared for the place names of the source of students. Beijing has been encoded as "8290", Beijing's municipal district has been encoded as "8291", and Dongcheng District of Beijing has been encoded as "8292". There are 3374 codes which correspond to the place names of the source of all students. The preparation of a unified code provides convenience for the data management. But the input information must be very accurate in order to find the right match code when there is new information needs to be entered into the system. An issue is an important subject that how to find the corresponding accurate information. The general solution is:

- (1). Drop-down list is used as the number of records is small, and the corresponding item can be selected by the user.
- (2). A table of standardized code is available to users when the number of records is large. The user will find the corresponding code from the table and write it to the text field.

The two methods of above are not to meet the principle of fast and accurate entry of data when the number of records is large.

2.2 Solution

The alphabet search is applied to the data entry process. For example, there is a student from Dongcheng District of Beijing,

Only parts of the alphabet in words need to be entered such as "bjd". All the records which contain these characters will be showed up by the application of fuzzy query technology .And then the user selected the corresponding option. There are three reasons for doing so:

- (1). to avoid inaccurate information appears. For example, Beijing has been encoded as "8290" in the database, while Beijing municipal districts corresponding code is "8291", if the record information is incomplete, the corresponding error code will be input to the database.

(2). to avoid homonym typos; Pinyin input method is widely used when Chinese characters are entered. Error of the homonym different word often appeared. If the wrong Chinese characters are selected, the correct data to the corresponding can not be queried.

(3). to reduce the number of character input. Chinese characters do not need to be entered therefore Phonetic retrieval will reduce the number of character input.

In order to quickly and accurately find the corresponding code, phonetic retrieval of information is used with Ajax technology in the process of system information input. Using the XMLHttpRequest object, data in the foreground is usually transferred to the background. Background transmits the results back to front after receiving the request with no page post-back. The entire process of data query and entry is in the same page. Users feel the page is efficient and sustained without refresh.

3 Efficient Fuzzy Query Based on Ajax

3.1 Fuzzy Query Methods

Fuzzy query methods currently used mainly in the following two:

(1). Fuzzy query in SQL

SQL [2] (Structured Query Language) is the most widely used structured query language. LIKE is a keyword that is used in the WHERE clause. Basically, LIKE allows you to do a search based on a pattern rather than specifying exactly what is desired (as in IN) or spell out a range (as in BETWEEN).The syntax is as follows:

```
SELECT column_name FROM [table_name] WHERE column_name LIKE  
PATTERN
```

PATTERN often consists of any combination of the four wildcards.

(2). the regular expression

Regular expression [3,4] is the special language for pattern matching. Regular expressions have the flexibility to construct a dedicated language to implement the pattern matching. A regular expression (regex or regexp for short) is a special text string for describing a search pattern. You can think of regular expressions as wildcards on steroids. The substring of the string in line with the regular expression pattern will be found and replaced.

3.2 Fuzzy Problems Encountered in the Query

Data is frequently invoked, each time the query is related to the large amounts of data in the table; fuzzy matching has encountered a problem:

(1). the amount of data is too large. Slow response times when the information is loaded via an Ajax request to the query page, and even cause crashes;

(2). Fuzzy matching accuracy is not high, fuzzy matching query results to a larger data set, users can not quickly choose the information they need.

Each of the two kinds of fuzzy query methods described in the above is inefficient when a large number of pending query data in the database. Using every method alone can not solve the problem of efficient and accurate query.

3.3 Fuzzy Query Algorithm in System

Pinyin of the information you want to query will be typed into Pinyin phonetic search box and the contents of the input is named as String1, the corresponding value in the table is defined as String2. The algorithm is as follows:

- (1). the string String1 is decomposed into a number of valid characters such as S1, S2 ... Sn;
- (2). the records are found which include one of the characters that are decomposed from the string String1;
- (3). Data in accordance with S1, S2 will be screened and defined as a fuzzy set A, and screened in accordance with the S1, S2, S3, sort the data in the fuzzy set A, defined as a fuzzy set B, and then find in accordance with the S1, S2, S3, S4, and so on, and finally obtained in accordance with the S1, S2,... Sn sort of data collection followed in the collection B;
- (4). Data sets in accordance with the S1, S2 ... Sn returned to the user;
- (5). User finds the accurate data in the last match of the data set to needed, and then record the data corresponding to the code is written back to the database.

In the system, filter out the data the String1 decomposition after the single character in the string String2, this process fuzzy queries using the SQL LIKE, completed by the database server; initial query results and then the order of the letters appear fuzzy matching, this process uses regular expression matching, done by the application server side, so that two fuzzy queries, respectively, sharing in the database server and application server. Thus the query efficiency is improved.

3.4 Realization of Fuzzy Query in System

Frequent calls to data are needed in the system, each time the query relate to a large amount of data in tables, while the fuzzy query results to the maximum extent close to the value of user needs, in order to facilitate the user to select. For each of query methods described in the previous, efficiency is low when a large number of pending query data in the database. Either method can not solve the problem of query precision and query efficiency. In order to solve this problem, two fuzzy queries are used. When the client query request is sent to the application server, application server accepts the request and submits SQL queries to the database server. After database server receives the request the initial fuzzy query will be performed. Then the query results are returned to the application server. In application server side the regular expression is used on the second fuzzy query thus matching accuracy of such data is relatively high. The two queries occurred in the database server side and application server side, sharing the pressure of either of servers; thence the speed of query is improved. The concrete realization of the process shown in Fig.1:

For example, a user want to query the string contains "beijing", the tables in the database contain 6937 records. The number of records returned from SQL fuzzy query is 811. And 86 records that match the regular expression are returned in the second fuzzy query. Thus the efficiency of query is improved greatly. User can get he most desired results through this query. At the same time the result can also match with the code in the database table. Information collection is completed efficiently.

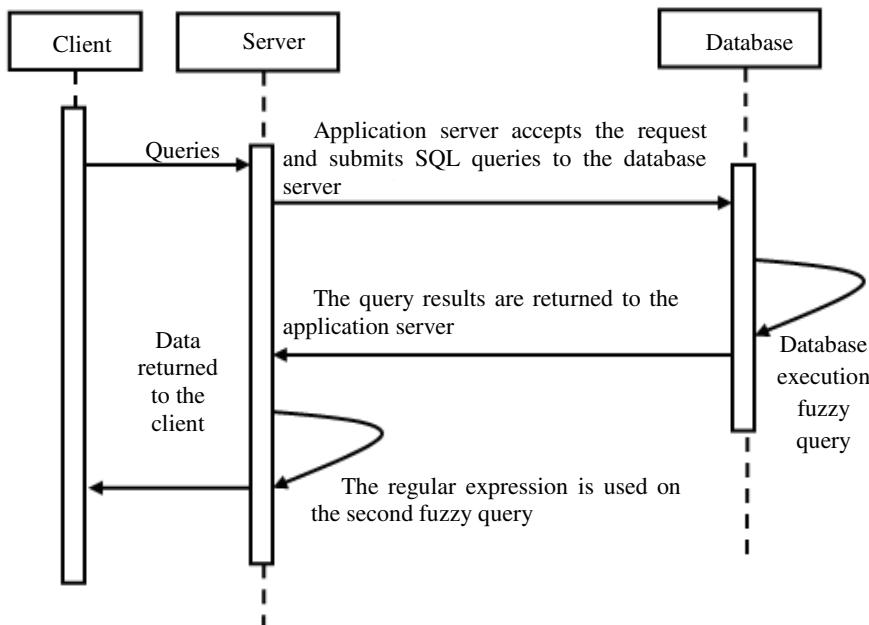


Fig. 1. Query process diagram in B / S mode

4 Conclusion and Discussion

Focused on solving the standardized code system, accurate data entry, dual fuzzy matching improves the speed and accuracy of query. In order to verify the availability of the method and accuracy of the query, program tracking is applied and the accuracy of query being tested for the module. Experimental results show that querying data with phonetic search is convenient for users; accuracy of the data collected is improved. The efficiency and accuracy of query are taken into account in the double fuzzy matching method to achieve the desired effect of the system.

Acknowledgement. We give the most heartfelt thanks to the reviewers for their great contributions on the early manuscript.

References

1. Ajax Tutorial, <http://www.w3schools.com/Ajax/Default.Asp>
2. Matter, R.: AJAX Crawl: Making AJAX Applications Searchable (2008)
3. Wang, K., Mo, X., Wang, W.: Phonetic Retrieval for Chinese Information in Database. Journal of Nanjing Normal University, 76–78 (2009) (in Chinese)
4. Liu, S.: Regular expression of Web data extraction study. Journal of Computer Programming Skills and Maintenance, 89–91 (2009) (in Chinese)
5. Yu, S., Zhou, S.: Application of Regex in Auto-Checking Paper of Programs. Journal of Computer Technology and Development, 224–246 (2008) (in Chinese)
6. Yang, C.: Fuzzy Query and Data Validations Based on Regular Expression. Journal of Computer Knowledge & Technology, 11–412 (2008) (in Chinese)
7. Hui, Y., Zhang, Y.-C.: Design and development of Ajax based experiment teaching platform. Journal of Experimental Technology and Management 28(1), 56–59 (2011) (in Chinese)
8. Zhou, R.-Q.: The Fuzzy Query Technology of Database Management System. Journal of Hainan Normal University (Natural Science) 22(2) (2009) (in Chinese)
9. Yang, C.-K.: FuzzyQueryandDataValidations Based on Regular Expression. Journal of Computer Knowledge and Technology 4(2), 411–412 (2008) (in Chinese)
10. Sipser, M.: Introduction to the Theory of Computation, pp. 2–50. China Machine Press, Beijing (2000)
11. Hopcroft, J.E., Motwani, R., Ullman, J.D.: Introduction to Automata Theory, Languages, and Computation, 2nd edn., pp. 1–112. China Machine Press, Beijing (2004)

Basin Hydrometeorology Real-Time Forecast System

Tao Peng¹, Cheng-zhi Wei², Jin-tao Ye¹, Jun-chao Wang¹, and Zhi-yuan Yin

¹ Hubei Key Laboratory for Heavy Rain Monitoring and Warning Research,
Institute of Heavy Rain, CMA, Wuhan 430074, China

² Jinhua Zumo Information Technology Co., Ltd., ZheJiang,
Jinhua, 321000, China
pt_mail@sohu.com

Abstract. The flood disaster has been threatens one of the human survival and development most serious natural disasters. The flood prevention situation is intense in flood season every year in China. The high technique content hydrometeorology pattern is needed urgently to develop the hydrology real-time forecast and the flood prevention decision-making service, sharpened the defense flood ability. Based on the real-time hydrometeorology monitoring network, the quantitative precipitation estimation (QPE), the quantitative precipitation forecast (QPF), the real-time flood forecasting technology, the basin hydrometeorology real-time forecast system apply the radar measure technology, the mesoscale rainstorm forecast pattern technology to gain the high space and time resolution hydrological data (rainfall and so on) and input hydrology model to carry on the real-time hydrometeorology forecast. Nowadays the system has carried on the hydrometeorology real-time forecast experiment and service on the Dan Jiangkou, Hanjiang River, Qingjiang, and made some achievements.

Keywords: Basin, QPE, QPF, Flood Forecast, Web-system.

1 Introduction

The flood disaster caused by the heavy rain has become one threat to human survival and development of the most serious natural disasters. According to statistics, the direct economic loss of billions of dollars is caused by the flood in our country every year¹. Since 1950 the previous floods in Yangtze River Basin (such as 1954, 1969, 1996, 1998, 2003, 2007, and 2010) brought huge loss to us. According to incomplete statistics, China has about 86000 reservoirs, the flood prevention situation is intense in flood season every year, and many basins, reservoir, lake and other urgent need the high technological contents hydrological and meteorological model to carry out real-time hydrological forecasting and flood control decision making service, improve the flood control capacity. Nowadays some objective condition is quietly changing, such as the integrated meteorological observation ability is obvious enhancing, numerical prediction ability is progressively getting perfect, the level of quantitative precipitation estimation is increasing year by year, hydrological model technology is innovating , geographical information technology is maturing gradually, network

communications and computer technology are also developing rapidly, they provided a very strong foundation and development opportunities for our country flood forecasting and warning technology development Basin hydro-meteorological real time forecasting system is developed to adapt to development and needs.

Based on GIS technology support, real-time hydrological and meteorological monitoring network, quantitative precipitation estimation(QPE), quantitative precipitation forecasting (QPF), real-time flood forecasting technology, basin hydro-meteorological real time forecasting system make full use of existing information of precipitation, to achieve the real time hydrological and meteorological forecast in the same platform, in order to provide technical support for improving the basin rainstorm flood forecast ability. The system has been successfully built system platform based on web, and carry out flood test and service, on the Han River in Danjiangkou, Wangjia ba in Huaihe, Shuibuya dam in Qing River, Zhanghe reservoir basin. and made some achievements.

2 The Structure and Design of System

The research thought of the system is to construct the hydrological meteorological forecasting system platform based on weather monitoring, forecasting, flood forecasting and GIS technology, the target is that the completed system can browse and query hydrological and meteorological monitoring and forecasting information in Internet network platform. The three layer structure (Browser/Client/Server) is used for overall logic structure of system, namely the rainfall monitoring, radar quantitative precipitation estimation, numerical forecast precipitation, flood forecast, hydrology and other product information are processed on Client/Server (C/S) system, the

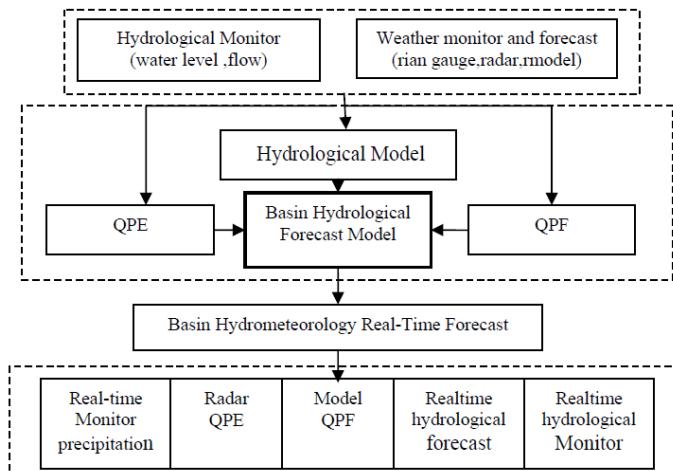


Fig. 1. The system structure and data flow

Browser/Server (B/S)system can complete the display of product information. The ‘Practical, real-time, advanced, extension, open, standard’ principle is adopted during the System development [2-3].

The system server runs on Microsoft WindowXP operating system, the Asp.Net4.0 is used to exploit system webpage platform, C # programming language is used, Microsoft SQL Server is used for webpage platform database, and connected through the ODBC database, Microsoft IIS5.0 is used for WEB server, IE browser is used for the client to access system.

System in the C / S level firstly obtain the basin real-time and forecast precipitation information by processing the rainfall station monitoring, radar detection, model forecast data, and make the corresponding graphics products, and then translate them into hydrological model needed information, input hydrological model to make real time flood forecast, finally the system in B / S level release the forecast information, graphic product through web, system structure and data flow can be seen in Fig 1.

3 The Exploit and Realization of System

3.1 The Pretreatment for Basin Basis Geography Information

The basin is a natural area, takes water flow as basis, the river as main line, the watershed as the boundary. Basin, its watershed, area, shape, direction and other basic properties are also different. The primary task for hydro-meteorological forecast system, that takes basin as the research object, is to obtain and process basin boundary, water and other basic information. The specific methods are as follows: firstly, Using the GIS technology to define the basin boundary and extract the river system based on the DEM data, and then to define the place and water in basin, finally to complete the digital work of basin basic information (the example is shown in Fig2).

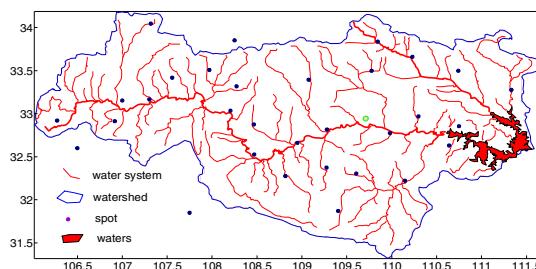


Fig. 2. The sketch map of dan jiang-kou basin

3.2 The Production and Transformation for Basin Precipitation Information

Precipitation is one of the major elements for basin, it is important for system research and exploit, that takes the hydrometeorology real-time forecast as key core, to how to

obtain a series of basin rainfall information, such as rain-gauge monitor, radar estimation, numerical model forecast, and transfer them into hydrological model input. During the system exploit, the following steps are complied to extract precipitation information products and design the interface between rainfall information and hydrological model:

Firstly, to know and be familiar with the time, storage, data structure of QPE, QPF, actual monitor precipitation information.

Secondly, the system adopt the some clipping algorithm⁴ and VC or VB program language to exploit the program to obtain the QPE, QPF, actual monitor precipitation information in basin.

Thirdly, the system adopt the Kriging interpolation algorithms⁵ and combine with the graphics system to generate basin 1h,12h,24h actual monitor,1h radar estimation, 60h accumulation, 3h, 6h, 12h, 24h period precipitation field graphics products automatically, and storage them specified server directory , and make the WEB server call graphics products.

Finally, according to the precipitation information input mode of selected hydrological model the system transfer the actual observation, QPE and QPF precipitation information, and input into the hydrological model.

3.3 The Construction for Basin Hydrological Forecast Model and the Realization of Hydrology Monitor Display

The basin locates in different locations, the climate characteristics have obvious differences, the first task for constructing basin hydrological forecast model is to select and determine the appropriate hydrological model (for example three water Xin'An hydrological model is suitable for humid and semi-humid region) based on basin geography climate characteristics [6-8], secondly the basin hydrological model needed to be established, finally it is needed for calibrating hydrological model parameters based on the known hydrological and meteorological data, and then obtain basin hydrological parameters.

During the real time forecast, the system start the hydrological model to call the precipitation information form the rain gauge, radar rainfall, meso-scale model forecast, to make the real time flood forecast, form the data file, and store them specified basin directory for database calling.

It is important for basin disaster prevention and reduction to make the hydrological monitor. The real-time hydrological monitoring data come from water network of the Ministry of water resources.

3.4 The Realization for Web Display Platform

The Asp.Net(C#) 4.0 are adopted to develop system display platform, Specific steps as follows:

The Display for Basin Overview: According to the selected basin, the Asp.Net database control search for basin information in the database server basin property list, and display Basin Overview through graphic form.

The Display for graphics products of basin actual monitor, radar QPE, model QPF: the system firstly built the control for basin actual monitor, radar QPE, model QPF display, and then design how to call the products in C/S level, and then construct the JS function in B/S level, finally the JS function will be called in the foreground or background to excite click events, and access graphic file in C / S level, and make it display on the B/S level.

The Display for Hydrological Forecast and Monitor Products: ASP.NET is good at Web programming, does not support chat and table, so the third control (such as VB MS-Chart controls, Microsoft.NET Framework, FLASH) needed to be adopted. During the system exploitation, we firstly adopt the ASP.NET to visit the hydrological forecast and monitoring data in database, and then use the FLASH control to display for hydrological forecast and monitor products.

4 The Forecast Experiment and Application Analysis

Since the flood season in 2010, the system make the forecast experiment on Hanjiang River, Dan Jiangkou, Qingjiang, Huahe Wangjiaba, Zhang-he basin, and made the good progress. In here, the brief introduction and analysis will be made based on experiment on Dan Jiangkou dam Han Jiang basin.

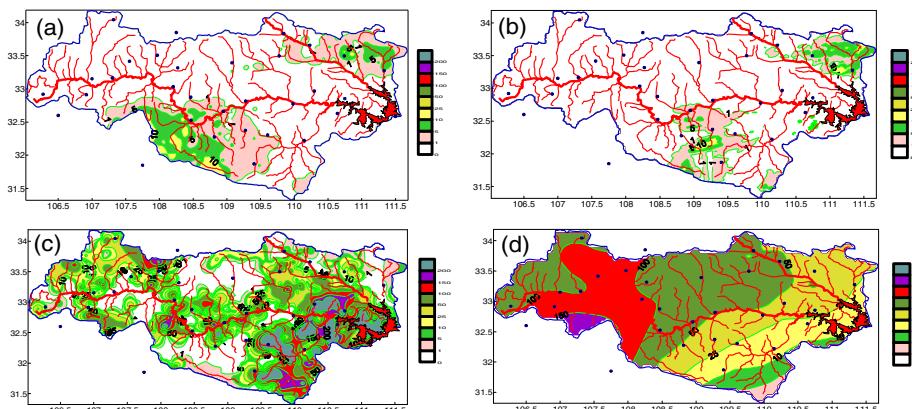


Fig. 3. Basin actual monitor and forecast precipitation (2010071803) (a) 1h actual monitor (b) 1h radar QPE (c) 24h accumulation actual monitor (d) 60h model forecast

Since the middle of July in 2010, the strong precipitation happen in the Hanjiang River basin, the max precipitation exceed the 500 mm, it leads to excessive flood in upper reaches of Hanjiang River water, the flood prevention situation become very serious.

Since the beginning of July 15, 2010, the system carry out the real-time forecasting experiments, make the basin hydrological forecast in the next 72 hours two times a

day, and provide hourly radar QPE, hourly monitor, 24h accumulation, 60 hour model forecast precipitation (Figure3).

Based on precipitation information of the rainfall monitor and model forecast, the system start the hydrological model to make the hydrological forecast, the result at 3pm,2010,7,18 show that the flood peak ($21442\text{m}^3/\text{s}$) will arrive in the Danjiangkou reservoir at 19 am,2010,7,19. Based on the hydrological monitoring information, it is can be found that the forecast result is more accurate, based on the comparison result at 8pm, 2010,7,19 (observation $19400\text{ m}^3/\text{s}$; forecast $20130\text{ m}^3/\text{s}$, seen in Fig4).

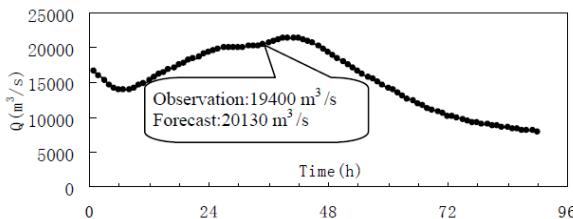


Fig. 4. The result of hydrological forecast

5 Conclusion

It is a significant task for basin disaster prevention and reduction to doing basin hydro-meteorological forecast service well. Basin hydro-meteorological forecast is complex system engineering, the system is confronted with new information carrier, it is an attempt for the author. During the process of system development, we make full use of new technology, new method in modern meteorological service to obtain actual monitor, radar QPE, model forecast multi-source precipitation information, and combined with hydrological model, and make the basin hydro-meteorological real-time forecast, make some intuitive graphical products on basin actual monitor, radar estimation, model forecast precipitation field and hydrology monitor and forecast curve, and provide support for flood control decision.

Form the flood season in 2010, the system makes the hydro-meteorological real-time forecast test on Danjiangkou Dam in Han River, and made some achievements, but at the same time some problems have released during the test. The basin hydro-meteorological forecasting system, still need further to be tested, and constantly updated and improved.

Acknowledgments. This work has been supported by the Special Research Project (Meteorology) of the Ministry of Science and Technology of China(GYHY200806002, GYHY201206028); The development of science and Technology Fund Project of Central China Regional Meteorological Center, CMA(QY-T-201202).

References

1. Bian, J., Li, S.L., He, J.H.: Risk Assessment of Flood Disaster in the Mid-lower Reaches of the Yangtze. *Journal of Applied Meterorological Science* 22(5), 604–611 (2011)
2. Zhang, Y., Xiao, W., He, W., Sun, Z.: Meteorological Information Publishing Platform Based on ArcGIS Server and VML Technique. *Journal of Applied Meterorological Science* 22(4), 498–504 (2011)
3. Hu, J., Liu, Z.: Development and Application of Flash Forecasting and Warning System for Medium-and Small-size River Basins. *Hydrology* 31(3), 18–21 (2011)
4. Liu, Y.K., Gao-yun, Huang, Y.-Q.: An Efficient Algorithm for Polygon Clipping. *Journal of Software* 14(4), 845–856 (2003)
5. Wang, Y.: Development and Application of a Class Library of Contour-Related Algorithms. *Meteorological Science and Technology* 38(4), 478–483 (2010)
6. Zhao, R.J.: Hydrological Simulation in Basins. WaterPower Press, Beijing (1984)
7. Peng, T., Song, X.Y., Shen, T.Y.: The application of radar quantitative precipitation estimation in flood forecasting during flood season. In: Proceedings 4th International Yellow River Forum on Ecological Civilization and River Ethics, pp. 309–314. Yellow River Conservancy Press, Zhengzhou (2010)
8. Peng-tao, Song, X.-Y., Yin, Z.-Y., et al.: Application of Radar QPE to Flood Forecast in Flood Season Hydrological Model. *Meteorological Monthly* 36(12), 50–55 (2010)

Internet Health Resources Development Based on Geographic Information System

Yun-Peng Zhang¹, Dan Li¹, and Ying Zhao²

¹ Jitang College of Hebei United University, Hebei Tangshan, China

² Library of Hebei United University, Hebei Tangshan, China

Yunpengzy@yahoo.com.cn

Abstract. This research from the health resources in terms of demand angle, analyse according to the allocation of resources situation, and configuration of the problems that exist, use of internet technology and to establish a collection and dissemination of health resources information network platforms. For the formulation of regional health planning and optimizing the allocation of resources may be provided the reference.

Keywords: Health Resources, GIS, Internet.

1 Introduction

In a certain social and economic conditions, health resources are the general term of state, society and individual to health department of comprehensive investment. It includes health human resources, health material resources, health financial resources and health information resources, etc[1]. Health resources are the basis of health department providing health service for social and the crowd. In health service activities, health services are the basic conditions[2]. And the state council puts forward the Chinese public health information network construction of the strategic policy of "overall planning, national leading, uniform standard, joint construction, the inter-communication and resource sharing". Among them the inter-communication and resource sharing is the ultimate expression of the spirit of the Internet[3].

2 The Significance and the Development Tendency

From a public health perspective, establish a unified information platform and data query platform. Realize data extraction, data management standardization, automation statistical analysis of diversification[4]. Integration with other systems collects statistical data, so as to share information resources[5].

From the perspective of health resources, by the end of 2010, the national health care institutions amounted to 0.937million. The national medical and health institutions have 4.787 million beds. And the national health personnel is a total of 8.208 million. The national total health expenses are expected to reach 1.9603 trillion yuan, and the per capita health cost 1440.3 yuan[6]. But in our country, some large and medium cities

don't make the best use of the advantage resources. If a map can directly reflect the distribution of health resources related information, people can easily understand health data of the actual distribution.

From the computer and GIS technology perspective, in recent years, the network technology and electronic map have developed rapidly. And geographic information system occupies an important place in the decision-making consultation in developed countries[7].

3 Materials and Methods

3.1 Function Analysis

A health resources distribution map to get application must establish a relatively perfect network information system to simulate the practical application of this design[8]. In the home page, users can browse the topographic map and zoom map at will. The project editor page should be realized modification, add, and delete health resources information of the classification of the menu grade. Map filling pages will should be realized the analysis of the data processing, using maps fill algorithm automatically generate results figure. With these function, the health resources distribution platform is comparatively complete. Fig. 1 is the whole platform system structure.

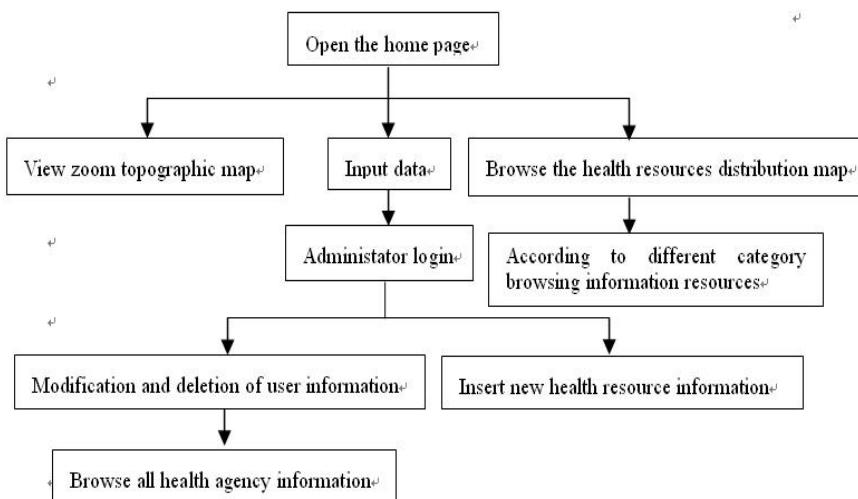


Fig. 1. System structure

3.2 Research Methods

Literature research: refer to the literatures, and study the relevant content, and access to relevant information.

Waterfall model method: w. Royce put this method the earliest in 1970. And it gives the survival of the activity cycle software fixed order, a stage of the activities after the completion of the transition to the next phase, finally obtained the development of the software product design method[9].

Prototype model method: prototype system is expected an executable version, which reflect the system properties of a selected subset. According to the purpose of using prototype different, prototype was divided into exploration, experimental and evolution model[10].

4 The Logic Design of the Database

The relevant health resources distribution map data table in the web site of the basic structure as follows.

Table 1. Menu list

No.	Field name	description	Type and length	Primarykey	For null
1	ItemID	Numbers	Nchar(20)	yes	no
2	UpID	Superior Numbers	nchar(20)	no	no
3	Number	Serial number	longint	no	no
4	Title	title	nchar(40)	no	no
5	Whos	category	nchar(50)	no	no
6	Clicks	times	longint	no	no

Table 2. Data table

No.	Field name	description	Type and length	Primary key	For null
1	ItemID	Numbers	nchar(20)	yes	no
2	ZooID	Numbers	int	no	no
3	Data	data	int	no	no

5 Part of the Design Code

5.1 Dynamic Menu Design

Menu control has two display modes: static model and dynamic model[11]. The static menu means that control is always fully extended. The entire structure is visible and users can click on any parts. In dynamic display of the menu, only the part is static[12]. And only the users' mouse cursor over the parent will only display its menu items. This system uses a dynamic model show menu.

The whole menu design the original code: Menu control source code

```
<asp:Menu ID="Menu1" runat="server" Height="18px"
style="font-size: large;" 
Width="126px"
DynamicHorizontalOffset="2" DynamicVerticalOffset="7"
Font-Names="Verdana"
Font-Size="0.8em" ForeColor="#990000"
Orientation="Horizontal"
StaticPopOutImageUrl("~/App_Map/Separator.gif"
StaticSubMenuIndent="10px"
MaximumDynamicDisplayLevels="4"

onmenuitemclick="Menu1_MenuItemClick">
    <StaticMenuItemStyle
HorizontalPadding="5px" VerticalPadding="2px" />
    <DynamicHoverStyle
BackColor="#990000" ForeColor="White" />
    <DynamicMenuItemStyle BackColor="#FFFBD6"
BorderColor="#996633" 
BorderStyle="Solid" />
    <DynamicSelectedStyle
BackColor="#FFCC66" />
    <DynamicMenuItemStyle
HorizontalPadding="5px" VerticalPadding="2px" />
    <StaticHoverStyle ForeColor="Maroon" />
</Items>
    <asp:MenuItem Text="查看资源分布图"
" Value="0">
        </asp:MenuItem>
    </Items>
    <StaticItemTemplate>
        <%# Eval("Text") %>
    </StaticItemTemplate>
</asp:Menu>
```

5.2 Data Entry Design

Map editor pages create three provinces, cities and counties option using Radio Button List control, with three ListBox control store provinces and cities county name. According to the choice of counties input survey data, click on the right side of the "save data" button to save the input data.

Specific event codeas follows:

```
protected void
RadioButtonList1_SelectedIndexChanged(object sender,
EventArgs e)
{
    ListBox1.SelectedIndex = -1;
    ListBox2.SelectedIndex = -1;
```

```

    ListBox3.SelectedIndex = -1;
    switch(RadioButtonList1.SelectedIndex)
    {   case 0:  ListBox2.Visible = false;
HiddenFieldZooSelected.Value = "000000"; ListBox3.Visible =
false; break;
        case 1:  ListBox2.Visible = true;
HiddenFieldZooSelected.Value = ListBox1.SelectedValue;
ListBox3.Visible = false; break;
        case 2:  ListBox2.Visible = true;
HiddenFieldZooSelected.Value = ListBox2.SelectedValue;
            break; } }
protected void ListBox1_SelectedIndexChanged(object sender,
EventArgs e)
{ HiddenFieldZooSelected.Value =
ListBox1.SelectedValue; }
protected void ListBox2_SelectedIndexChanged(object
sender, EventArgs e)
{ HiddenFieldZooSelected.Value =
ListBox2.SelectedValue; }
protected void GridView1_RowDataBound(object sender,
GridViewRowEventArgs e)
{ if (GridView1.Rows.Count > 0)
    { ListBox3.Visible = false;
        PanelImportant.Visible = true; }
    else
    { ListBox3.Visible = true;
        PanelImportant.Visible = false; } }

```

6 Summary

To sum up, this design can make people know the distribution of health resources intuitively, quickly and timely. It is convenient for health department leaders to undertake unity and the dynamic regional health resource planning. This design can improve the management level of decision-making. And finally realize the reasonable configuration talent, material, and other resources.

References

1. He, N., Xin, Y.: Tianjin health resource allocation situation and equity research. Chinese Health Service Management (3), 174–176 (2010)
2. Lv, K., Ma, Y.: The current Chinese medical health resource allocation is the main problem. Chinese Disease Control Magazine 12, 626–628 (2008)
3. Luo, W., Yu, Z.: The Internet technology in public health impact analysis of information platform. Sichuan Health Management Cadre Institute 27, 141–142 (2008)
4. He, Z.: Health resources reasonable use is the key of “medical reform” to success. China Pharmacy (22), 2017–2019 (2010)
5. Wen, J., Jiang, B., Li, Z.: Medical health resource utilization of our country research progress. Chinese Journal of Coal Industry Medicine (7), 1103–1104 (2008)

6. The health ministry office about the need to 2009, national health statistics information work notice (who do ensemble is sent, 29) (2009)
7. http://law.baidu.com/pages/chinalawinfo/11/71/6cfe6d00a79ea5d0e091ac1a146cf71e_0.html (February 2009)
8. In 2010, health service development statistical bulletin
9. <http://www.moh.gov.cn/publicfiles/business/htmlfiles/mohwsbwstjxxzx/s7967/201104/51512.htm>
10. Li, X., Shi, A., Zhang, G., et al.: Health resources space layout optimization model of the introduction. *Modern Preventive Medicine* 35, 1281–1283 (2008)
11. Peng, Z., He, J.: China health resource allocation of the present situation, the existence question and the emphasis of the reform of difficulty analysis. *International Medical Health Herald* (19), 21–23 (2005)
12. Luo, J., Wang, H., Cui, K.: Shanghai medical resource allocation condition analysis. *Chinese Journal of Health Statistics* (5), 466–469 (2009)
13. Song, Y., Li, X., Zhang, S.: The new rural cooperative medical problems and the sustainable development research in Hebei province. *Chinese Health Economics* (8), 34–36 (2009)
14. Zhang, Z., Zhang, S., Li, X.: The 2011-2015 in hebei province in health resource allocation standard formulation of ideas and methods. *Chinese Health Economics* (6), 40–42 (2010)

Research of Mobile Library Information Service

Yun-hong Lv

Hebei United University Library Jianshe Road 57 Tangshan Hebei China
Lvyunhong@sina.com.cn

Abstract. The article in the full investigation of mobile library in foreign countries and domestic development, and service mode of mobile library information based on. Analysed China's mobile library's development condition as well as the mobile information service situation of university library. The popularization of 3G mobile phone for mobile library information service infuse powerful vigor, promote mobile information service of library, and make library information resources available. At the end discussed mobile library service prospect.

Keywords: 3G, Mobile Library, Information Service.

1 Mobile Library Generation and Development

Mobile library as an information service of the digital library is a new service mode, which is relying on the mature wireless mobile network, internet and multimedia technology, so that people do not suffer time, place and space constraints, through the use of a variety of mobile devices (such as: mobile phone, personal digital assistant, notebook, electronic books, etc.). Convenient and flexible for library information query, browse and get a new library information service [1]. It opened up a new world for digital library information services. 3G environment of intelligent mobile phone has powerful function of data transmission, the library can be timely to short message, Multimedia Messaging Service, and push message (pushmessage) wait for a variety of ways to carry out the service, with any media can not match up, real-time, wide audience, more interactive as advantages.

1.1 Mobile Library Development in Foreign

Foreign mobile library can be traced back to around 2000. Japan and Europe in the mobile communication technology is more advanced. Japan Toyama University Library in September 2000 developed the I-mode (information mode) mobile phone library bibliographic query system. University of Tokyo Library in May 2001 opened I-mode mobile phone library bibliographic query system. Finland University of Helsinki library beginning in fall of 2001 using mobile phone short message service, in July 2001 South Sogang University launched mobile library through mobile phone. To

date, Finland, Japan, the United States, Britain, South Korea, Singapore and other countries have some libraries in trials with mobile information service. [2]

1.2 The Development of Domestic Mobile Library

Domestic mobile library although development relatively late. But some have established their own mobile library system. Especially since 2003, has been widely applied. Shanghai library opened our country's first real significance of the "mobile library", the reader can carry the mobile device, to enjoy the advisory service. Having opened the service projects include: the opening information, organized reader activity information, bibliographic retrieval, document request, reference, lecture reservation, and reader mail etc... And then the Liaoning library, Xiamen library, Shenzhen library also continued to open a "mobile library".

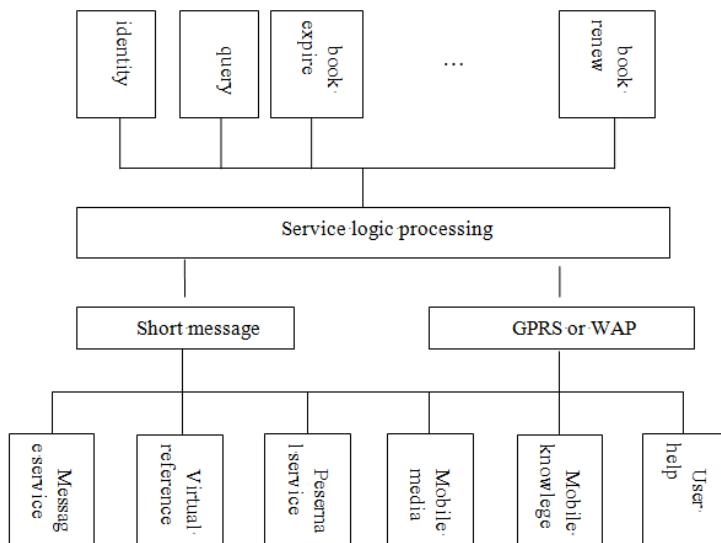


Fig. 1. Mobile library service process

"China mobile phone Internet User Research Report in 2010" survey data shows, in 2010 18-24 years old mobile internet users accounted for more than half, reached 57%; 25-30, netizens accounted for 23.2%, in the second; mobile phone Internet access frequency, "a daily average of mobile phone internet time" in 2010 the proportion has reached 88.8%, rose by 6.8 percentage points than in 2009, the internet will improve. From 2010 and 2009 survey data contrast, the age of 24 years and user group's ratio has increased substantially. And in the next three to five years, young users will continue to be main feature.

This online group is coincided with the college students. The data shows, to the mobile phone as the medium to develop mobile library information services, not only

has a huge readership, and conform to the development trend of mobile digital reading, but also fully utilize the library in the information communication in the value.

2 Mobile Library Information Service Mode

In the 3G communication technology, library use their mobile network and mobile phone information as a receiving platform, to provide users with a powerful information resources which will become an important new mobile library service mode.

2.1 The Wireless Network Service

WAP (Wireless Application Protocol) the mobile communication device has access to the internet and open international standard, is a kind of narrow bandwidth data communications protocol. It is the mobile library online information retrieval system. It uses a markup language WML (not HTML) which was defined as XML110, composed of a series of standardized protocol, used wireless communication device, can be used for the Internet access, including access to the WAP page on the site, email, chat communications, etc. [3]

Most of the current webpage are based on HTML (Hypertext Markup Language), this webpage for mobile phone internet users, not only slow, but also complicated to operate. WAP makes the mobile internet had a standard, its target is make internet rich information and advanced business into mobile phones and other wireless terminals. WAP definition universal platform, the internet online HTML language information into WML (Wireless Markup Language) to describe the information, displayed on the mobile phone display screen. WAP requires only the mobile phone and WAP proxy server support, and does not require the existing mobile communication network protocol for any change. The URL is for WAP.XXX.COM.

2.2 Streaming Media Service

In 3G network, whether it is CSD or GPRS. Their data transmission speed are not adequate to meet the broadband video on demand, streaming media business needs, and the 3G technology can fully meet the requirements of the business. The use of 3G high transmission speed, we can provide readers with more and more multimedia resources. Processing “sound, color and taste of the multimedia e-book,” to establish series of video lectures, multimedia database service. Apply of portable multimedia service system, readers can more easily at any time and place, enjoy more rich and colorful video information service than traditional text resources, and it is a step of library service of readers actively value. [4]

2.3 Video Reference Advisory Service

In 3G times visual telephone business is which communications industry focused on in China. It makes the communication more vivid, can make both sides for face-to-face communication, the service once introduced into the reference and consultation system

can make the reader whenever and wherever possible and the library advisory staff for face-to-face communication, even though there is a language barrier can also through gestures or picture for reference, makes the service more face to face. [5]

2.4 Personalized Customization Service

In the process of mobile phone media, mobile phone and mobile phone reading has become a bright spot. Library can learn from media launched mobile phone newspaper form of special service in the library, personalized service and dynamic information integration processing, according to the user's interest, or in response to a user request data for statistical analysis, transfer to the user through a remote server active text, sound, images, data and other information in a variety of formats. The library also provides partial copyright license the article download and browsing, enable the reader to have a mobile library and knowledge library [6].

Through personalized custom services, the reader may according to own demand through the mobile phone to customize their interested information or services, customized services include new information bulletin, an appointment to the books, borrow books due to remind, remind all kinds of information seminars. In order to obtain the services they want whenever and wherever.3domestic mobile library development.

3 The Development of Domestic Mobile Library

3.1 The Mobile Library Development Status

Beijing Shu-sheng company lasted three years on “mobile library” of the overall solution for development, hope that through technical means to help the library to achieve internal digital resources in the process of moving the unified retrieval and text reading, and make the library services to a new level. Shu-sheng solved many technical problems, such as electronic resources based on terminal equipment in full display and reading problems, different data storage format cannot be read through the questions, as well as various handheld devices to realize unified search problem. Peking University Library Director Zhu Qiang believes that, in new environment libraries will change. In June 2010 Shu-sheng launched the world's first mobile library solutions; in June 2011 he launched the world's first mobile library -- the mobile library, which also marks the library from the era of digital library into the mobile library. According to the president Jiang Haifeng of this company introduced, mobile library is a strong interactive carrier, its reading tool includes mobile phone and handheld reader, it makes the mobile library become a professional exchange platform, and associated with the search engines of different knowledge sources.

Although the mobile library in domestic colleges and universities application has not universal, but its development trend is irreversible, As the larger domestic literature management system Huiwen literature management system, has developed the mobile phone library system, which is to provide readers with mobile phone and other mobile terminal based on the library information service system. The reader can in any place,

any time through mobile phone short message and library system interaction. Huiwen mobile phone library system contains two products: 1, the short message service platform; 2, the mobile phone OPAC system. Covers the basic business library content, the reader need not go to the library can also enjoy the library service possible.

3.2 The Technical Problem in Mobile Library Development

3G portal founder Mr. Zhang Xiang-dong and his team created the “GO” browser, the appearance of “GO” filled the gap between WWW and WAP. WWW will no longer be a computer to browse the website, what makes 3G broadband life comes ahead of schedule, and will popularize and attract more new mobile phone Internet users. When the webpage browsed by “GO”, through a special page compression technology, will be more time-consuming pictures and text can be displayed separately, if need we can directly click on the text into the page, not need to spend time waiting, greatly reduces network traffic in a mobile phone access to the Internet, improve the speed at the same time, greatly saves the user flow of cost.

“GO” using in WAP web site, faster than the same industry browser 20%, while browsing the WWW on-line performance more surprise. The last mobile phone browser browsing WWW sites, mostly complete is a simple conversion, speed is slow and often can not display properly, typesetting confusion leads to poor user experience, and the emergence of “GO” complete these problems to a minimum. This important technical created favorable conditions for mobile library.

3.3 Current Situation of the Development of Mobile Library in Colleges and University

As shown in the following table: for some colleges and universities development on mobile information service investigation, which investigated ten for domestic famous universities [7].

Table 1. 2012 universities library mobile phone list (January 2012)

University name	Query path	Service mode	Service function
Peking University	Home page- Service - short message service	short message service	Library notice, literature due to remind, appointment to the book reminds and overdue books overdue reminders The circulation information notification (maturity, expired, overdue, appointment books etc.), personal information (borrowing, appointment, renew, cancel wait) library catalog query
	Home page- service short guide - short message message service	message service	The library bibliography query, personal borrow information query, electronic resources database, login / registration, help information
Tsinghua University	Home page-News- Mobile Phone Library System http://166.111.120.15 1:8080/thulib	WAP Service	Web

Table 1. (*continued*)

Fudan University	Home page-reader service-books borrow - reminder notice	E-mail remind	Maturity book, automatically send E-mail reminded to return
Zhejiang University	Home page -reader short service - Notification message Service	service	Book due to remind, book reservation remind
Shanghai Jiao Tong University	Home page - service guide -borrow service	short message service	Overdue books recall service
Nanjing University	Home page - e-mail reminder service	E-mail service	By sending e-mail access to library information
Wuhan University	Home page - Service information – E-Mail notification	E-Mail notification	Sending a book publication, electronic resources, new service projects and public information by E-mail notification system
Renmin University China	Home page - User of service - E-mail consulting	E-mail consulting	Through the E-mail consulting services
Jilin University	Home page- Virtual reference	Website	By visiting the website online to answer readers' Advisory
Zhongshan University	Home page - Borrow (E-mail) outside reading - appointment Make an appointment	() notice	Reservation books by e-mail

4 The Mobile Library Development

4.1 Establish Mobile Library Portal Website

To support the user conveniently and effectively collect information, and effectively use the information resources, mobile library should increase propagandist strength, extend knowledge scope and become bigger and stronger. This requires the library information resource deeply processing, establishing of standardized database, providing authoritative and reliable subject information navigation; strengthening the integration of digital resources, providing uniform retrieval platform and information service system; establishing an interactive intelligent digital reference service mechanism; creating rich and practical characteristic library, achieving “full information release” service mode, the librarian and user can conveniently pushes the information to a home database, simplified operation. Therefore, establishing library portal become a great step on the mobile library construction.

4.2 Constructing the Library Consortia

Mobile library construction based on intelligent mobile phone with 3G, certainly will need certain labor power, material resources, especially to a number of flexible use of WAP language programming technical staff support. If only rely on a library to

complete its quality must not high. This requires us to set up the concept of library, component library and between the library and the information organization alliance, jointly cope with the mobile library construction problems.

Strengthen cooperation of all types of libraries and other information agencies, and the province, the country or even in the global village concept to all over the world, the library as a social library to view, establish a "library" of the service concept, resource, technology, personnel, service sharing. The era of the 3G, requirement" library" and communications equipment manufacturers, operators, strengthen cooperation, unified standards and protocols, realize digital library mobile service "seamless" experience. [8].

4.3 Accelerate the Library 2.0, WAP2.0 And 3G Fusions and Development

3G application in library is the following of Lib2.0, service mode and service concept change again. The era of the 3G, to the user the most attractive is to enjoy a more personalized, more convenient mobile information service. Therefore, how to grasp the different requirements of users is the key of library realize new-style mobile information service. This concept is consistent with Lib2.0 taking the user as the center, taking web as the platform, emphasizing the interactions with users, focusing on user experience. Lib2.0 and WAP2.0, 3G integration, not only including RSS / ATOM, Blog, Instant Message (such as MSN, QQ, Skype etc.) and other commonly used techniques for fusion, also includes the Lib2.0 service to mobile phone transplant reproduction into a new information service, such as the mobile blog, community of readers, such as mobile IM[9]. The technology based on 3G service application, will lead the communication service function from the desktop to the handheld transition, and realize information of high portability.

In short, intelligent mobile phone with 3G relies on the establishment of mobile library, is available in Lib2.0 based on a qualitative leap, the cable network to solve information service can not overcome the difficulties, can quickly solve the body outside the library, not convenient to readers' information needs, so that the reader is very convenient access to library information resources service any time and any place outside school, truly ubiquitous, but also fully embodies the value that library as a document information center, and is bound to bring infinite vitality of library[10].

References

1. Ying, J.: Situation and Development Strategies of the Mobile Library Service in China. Library Construction 12, 75–78 (2011)
2. Sun, S.: Comparative Research on the Mobile Services of Libraries at Home and Abroad. Science and Technology Information Development and Economy 19(21), 63–65 (2009)
3. Zhao, Y.: Research and Design of the Mobile Library Based on WAP Technology. Library and information service 53(21), 106–109 (2009)
4. Huang, J.: 3G technology applied in the digital library. Information Research (4), 90–92 (2009)
5. Wang, Q.: China mobile library development foreground analysis in 3G area. Science and Technology Information (18), 332 (2009)

6. Wu, A.: Mobile Internet platform of mobile library services. *Journal of Library Science* 1, 80–83 (2012)
7. Yu, S., Ming, J.: Survey and Analysis of Mobile Information Services on Domestic University Mobile Libraries. *Library Journal* 91(9), 45–48 (2011)
8. Ma, A., Yang, G.: Investigation and Consideration of the Situation of Mobile Phone Service in University Libraries in China-Taking “211 Project” University as Example. *Library Work and Study* (12), 87–90 (2009)
9. Zhu, J.: Some Views on the Developing Opportunities and Challenges of the University Library from the Prosperity of Mobile Library. *Journal of Modern Information* (11), 138–141 (2011)
10. He, W.: On the Development Trend of the Mobile Library. *New Century Library* 41(1), 67–68 (2009)

Civil Explosives Distribution Management Information

Xue-yu Mi¹, Peng Zhang², Lin Zhang¹, and Bo Dong¹

¹ College of Civil and Architectural Engineering, Hebei United University,
NO. 46 Xinhua West Street, Tangshan 063009, Hebei Province, China

² Beijing Roads and Bridges Construction Group Corporation Limited,
NO. 318 South Garden Town New Village, Beijing 100068, China
mxy@heut.edu.cn, pengto@hotmail.com,
jlu.zhanglin@gmail.com, zhangyx0663@163.com

Abstract. The daily management of civilian explosives demands for computer systems is manifold. It requires the management of geographic information, processing documents, spreadsheets data, and depending on the circumstances of macro-and micro-technology management decisions. Research group has developed a test for civilian explosives distribution management system. The environment based on warehouse information. Upon completion of the basis of information processing environment, it is the transport vehicle monitoring, risk assessment and transport routes to optimize transport routes for the synthesis of organic whole. Distribution system model is based on Surper map-based information system, the core of the transport route optimization, risk assessment, emergency rescue for the content of the integrated, space-based model.

Keywords: Explosives, Distribution, Information Management.

1 Introduction

In recent years, with the growth of the demand for civilian explosives, the rampant terrorist activities, the distribution of explosives safety management system in place and the increasingly prominent role, warehousing and transportation and distribution date technology to keep explosives for civilian life and safety management vitality. As the special nature of explosives, distribution management with professional, demanding, in the event of an accident caused great harm to features. Thus, application of computer systems will update the traditional means of storage, transportation and distribution, specific work can be divided into two aspects, one is related to the management and geography, such as the choice of transport routes, vehicles and the impact of the explosion range of explosives transport vehicle location monitoring; on the other hand, it has nothing to do with the geography of management, such as explosives out of storage, inventory, ledger and other traditional management of daily business.

2 System Requirements Analysis

System Requirements analysis is a key step for the life cycle of the system .It is able to put the overall system functionality and performance of the system concept described as the specific needs of specifications, through the system requirements analysis. We can establish the basis for system development on this basis.

2.1 System Analysis Prepared

1) User analysis

The system is mainly for users of commercial explosives in the warehouse storage warehouse management staff, executive leadership, public security departments, and the enterprise security management, considering the current domestic civilian explosives safety management status. They have some experience in practical operation.

2) Technical Analysis

The sector enterprises are basically equipped with high performance computers and peripherals, with the development of computer technology and the rapidly growing popularity. Each level of equipment and technology departments can meet the needs of general application to run. Therefore, this system has a strong prospect, with the objective conditions of application.

2.2 Feasibility Analysis

1) The feasibility of data

Data is the basis for system operation, based on civil explosives warehouse and distribution management system, you need a more complete from the depots to get from the blasting site vector graphics. So as to form a set of query, analysis, maps, browse, locate, and other functions in one of the explosives warehouse and distribution management system. In Tangshan City, sijiaying vector mine as an example, a more complete road network, depots and other layers, so the system implementation, the data is feasible.

2) Technical feasibility

Using more current popular development language from Microsoft VC, with component-based development platform SuperMap Object for secondary development. Database using Microsoft SQL Server 2000 as the core database. It has a strong current, a good interactive interface and powerful spatial analysis capabilities.

3) Economic analysis

Based on super map distribution management system for civil explosives for explosives safety management and to provide the best possible logistics and distribution solutions. It saves distribution costs and improve distribution efficiency, reduce the intensity of work scheduling staff, and create more economic benefits for the mine.

3 Data Flow Design

3.1 Overview

Main data is based on the Tangshan City, szechuating vector map of mine and some explosives storage data. Specific data processing flow is as follows:

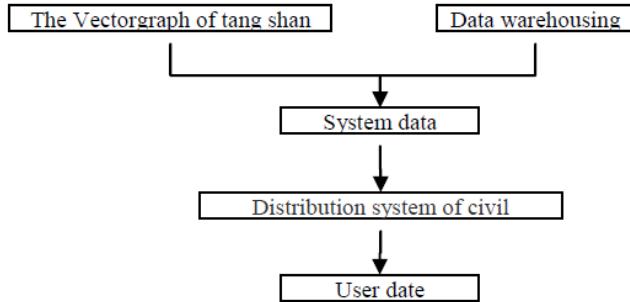


Fig. 1. The model of system data processing flow

3.2 System, the Logical Structure

System to express the information to the user, real-time, easily. System to map data, attribute data with the use of components and integrated development platform. System architecture shown in Figure 5.

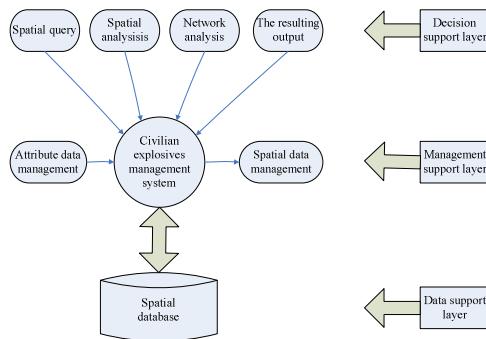


Fig. 2. The model of system logic structure

3.3 Data Design

Data is divided into three parts, one part of the layer of spatial data, in part, explosives storage information, and part of the video data.

1) spatial data

Data using Super Map work space and manage the SDB and SDD documents. Data space coordinate reference system for the Xian_1980_3_Degree_GK_CM_102E. In addition to the geometric network data in Road Net data sets inside, the other spatial data layers including, as shown in Table 1. Each dataset comes with the property sheet, used to record the various attributes of each element.

2) The attribute data

Attribute data is based on SQL Server 2000 set up a database called My Data. This database is primarily based on the data sheet. Attribute data of the system operation, mainly for the data table access operations. Attribute data, including depots, road data, and details of surface features around the path. The attribute data and video data using SQL Server database, a unified management. Which the video data stored video files relative storage path. Between spatial data and attribute data associated with a unique identification code by the Object ID.

Table 1. The layer data

Function	Data set name	Layer Name	Layer type
layers	Lu Deng	Street lamp	Point map layers
	Bi Lei	Lightning rod	Point map layers
	Kan	Kaner	Lline layer
	Wei Qiang	Wall	Line layer
	Dian Li Xian	Power line	Line layer
	He	River	Surface layer
	Fen Chang	Cemetery	Surface layer
	Lu	Road	Surface layer
	Fang Wu	Houses	Surface layer
	CaiShiChang	Quarry	Surface layer
	HanDi	Dry land	Surface layer
	Shui	Water	Surface layer
	CaiShaChang	Adoption battle	Surface layer
	CaiDi	Vegetable plot	Surface layer
	ShanDi	Mountain	Surface layer
	ZhuJi	Note	Note layer

Table 2. Explosive charge library of attribute data

Column name	Type	Explain
Object ID	Int	Low unique identification ID
Name	VarChar	Low signal name
Person in charge	VarChar	Medium person in charge name
Inventory	Int	Low contact phone
Phone	VarChar	Medium telephone number

4 Overall Design System Function

4.1 The Basic Function of the System

Through the computer to achieve the explosives distribution business information processing. Logistics and distribution analysis of the properties making use of GIS and spatial data integration, data management features. The system will accept business in front of the information collected collected, prepared, stored in a database. In the scheduling process in accordance with schedule of goods that need database information to generate the best scheduling programs to achieve the vehicle load, driving route of the unified plan. By Tangshan, szechuating mining explosives distribution management analysis, we have civilian explosives distribution management system's basic functions. It includes the delivery of optimal scheduling of explosives management, warehouse management of explosives, explosive library of video management and alarm management, vehicle scheduling source information management, transport routes, GPS positioning and risk analysis in five areas.

4.2 GPS Location Tracking Transportation Vehicles

Vehicle monitoring system consists of global satellite positioning system (GPS), communication systems, control centers, automotive equipment, and external interfaces. The basic function is to be achieved:

- 1) positioning of the blasting equipment, transport vehicles in real-time monitoring of the operation of the vehicle position, speed, driving directions, travel time data at the monitoring center can be displayed on the electronic map;
- 2) tracking the blasting equipment transport vehicles timely warning when unexpected events occur, the system automatically track police vehicles, police vehicles recorded before and after running;
- 3) cross-border warning function can be pre-set demolition equipment, transport vehicles to run routes, if the vehicle during operation deviates from the set route, the system will automatically warn the monitoring center to alert drivers and monitoring personnel to take immediate measures; a state of emergency, the vehicle system automatically send alarm signals to the monitoring center;
- 4) Alarm function, the monitoring center can be suspended by sending control commands to force a controlled vehicle;
- 5) The automatic locking function, the monitoring center can be suspended by sending control commands to force a controlled vehicles.

4.3 The Overall Design Flow System Function

Based on the above analysis of the overall system functions derived design. System, including inventory management, distribution and transportation management, path risk analysis, vehicle monitoring and control of the map. Inventory management including the storage management, inventory, the storage ledger and so on.

Transportation and distribution management information from the vehicle, receipt management, loading plan, route planning and vehicle rescue of five parts. Among them, the map control functions include, map zoom, full map display, map roaming, features in the query, the first window after window, and calculating the angle, length and calculating, data updates, flowcharts. The GPS function of the transport vehicle control modules, mainly for transport vehicles in-transit monitoring, as well as unforeseen circumstances and re-arrange vehicles to achieve real-time vehicle scheduling information. Path analysis and evaluation module, the main transport vehicle once the explosion, the impact on the surrounding environment, the scope of the search and the surrounding nearest hospital, in order to improve rescue capabilities. Specifically shown in Figure 3~8.

4.4 System Characteristics

Mine safety explosives warehouse and distribution management system with the following characteristics:

- 1) The acquisition, management, analysis and output a variety of security geospatial information;
- 2) information management, analysis and output of spatial and dynamic nature;
- 3) In order to achieve the distribution of mining explosives transportation safety, transportation, vehicle scheduling purposes, with the regional spatial analysis and dynamic analysis of multi-element integrated predictive power;
- 4) Analog transportation of explosives and explosion of the sphere of influence;
- 5) The graphics and text mode output;
- 6) to build a civilian explosives inventory data information systems and transportation of explosives combined distribution line information network platform.

5 Implementation of the Main Interfaces and Functions

1) buffer to achieve: the use of Buffer function, the function of the design ideas are as follows:

```
/ / Start:  
/ / Find the point;  
/ / If (point does not exist)  
/ / Return;  
/ / Otherwise  
/ / Positioning point position;  
/ / Point to another buffer layer and the cross-point  
operations;  
/ / If (result set is empty);  
/ / Return;  
/ / Otherwise;  
/ / Output results;
```

```

/ / Generate the geometry;
/ / Fill the buffer range of geometry;
/ / Graphic display;
/ / End;

```

2) Shortest Path: The main advantage of ShortPath function, basic design ideas are as follows:

```

/ / Start:
/ / Find the two points;
/ / If the
/ / One thing or two full of empty space;
/ / Return;
/ / Otherwise
/ / First point and line layers intersection operations;
/ / Second point and line layers intersection operations;
/ / Iteration;
/ / Find the shortest path;
/ / End;

```

3) Other operation of the map interface is mainly the ICommand interface, create thematic maps of the

Show Theme function, and operation of the database ADO.NET interfaces, and GPS simulationfunction Displaysattlite, this paper will not go into the concrete realization of these functions.

6 Conclusion

In research and development, involving geographic information system technology, relational databases, high-level computer programming language, and other advanced computer technology. System for civilian explosives safety management application of modern technology and new achievements in science and technology, modern management to improve the level of this work to explore a new path, creating social and economic benefits. Appropriate expansion of the system can also be used in a number of related areas, such as: flammable materials storage, dangerous goods transportation, public security command, firefighting, civil explosive company goods warehousing, transportation routing, risk assessment and other relevant departments. It's also close with environmental protection, network planning, vehicle scheduling and other fields. It will has a good prospect.

Acknowledgement. This work was supported by natural science foundation of Hebei United University (z201102), natural science foundation of Hebei province (F2010000976) and Study and Development of Science and Technology of Tangshan (10110206C).

References

1. Xiao, J.H.Z.Y.: Potential Safety Hazard and Countermeasures of Civil Explosives Road Transportation and management in China. *Engineering Blasting* 16(1), 86–88 (2010)
2. Wang, T., Zhang, S.B., Li, B., et al.: Safety situation survey of small civilian explosives repository on the ground in the city. *Journal of Safety Science and Technology* 6(2), 109–112 (2010)
3. Hu, J.: Considerations on the Safety Management of China Civil Explosives. *Shanghai Higher College Transport* 20(2), 50–53 (2010)
4. Liu, Y.X., Zheng, Z.Y.: Civil Explosives Safety Management and Countermeasures. *Mining Technology* 7(3), 153–155 (2007)
5. Zhang, Y.Z.: Blasting equipment operation and management. Metallurgical Industry Press, China (2004)
6. Arturo, H.A., Salvador, B.R., Carlos, A.C.C., et al.: Handling Constraints using Multi-objective Optimization Concepts. *International Journal for Numerical Methods in Engineering* 59(15), 1989–2017 (2004)
7. Butler, M., Herl, I.P., Keenan, P.B.: Integrating information technology and operational research in the management of milk collection. *Journal of Food Engineering* 70(3), 341–349 (2005)
8. Roberto, S.D.C., Barbara, M.: Risk analysis for road and rail transport of hazardous materials: a GIS approach. *Journal of Loss Prevention in the Process Industries* 17(6), 483–488 (2004)
9. Wu, H.B., Yan, S.L., Qiu, X.L., et al.: Programming of Civil Explosives Information Management System. *Blasting* 22(4), 127–129 (2005)
10. Wang, C.P., Yu, L.S., Ren, S.H.: Design of the Blasting Parameters in Tunnel Excavation Based on Visual Basic. *Journal of Northwest Hydroelectric Power* 23(1), 81–83 (2007)
11. Xia, J., Zhou, M., Zhou, X.G.: Analysis of Destroy Accident of Explosive Remnants of War Blasting. *Blasting* 49(1), 117–121 (2012)
12. Fang, J., Shi, Y., Wang, L.L.: Design and Implementation of the Remote Controlled Acquisition System of Explosive Ordnance Disposal Robot. *Microcomputer Information* 26(10), 163–165 (2010)

Research of the Pension Insurance Information Management System Based on the J2EE

Ying Chen¹, Xiaosi Xu², Jun Zheng³, and Hong Wang³

¹ Department of Student Affairs, Hebei United University, Tangshan, China

². Organization Department, Hebei United University, Tangshan 063009, China

³ Finance Department, Hebei United University, Hebei Tangshan, 063000, China

Abstract. With the increasing requirements of the support on the pension system technology in current socio-economic situation, the traditional endowment insurance management services means and conventional statistical system are increasingly difficult to meet the needs of the development under the new situation. Therefore, to establish an appropriate scale, in line with the pension insurance system in Hebei Province, we can effectively solve the data gathering integration, dig analysis, play an overall collaborative performance and other issues, also significantly improve and enhance the management means and service level of the pension insurance in Hebei Province. In this article, fully base on the comparison and analysis of the current software platform architecture, J2EE technology provide systematic software architecture, systematic formation of function modules, and technologies routes realized by system. It provides the support to establish a standardized system of pension insurance business, scientific macro-monitoring system and the complete public service system.

Keywords: Pension insurance, Information management system, J2EE.

1 Introduction

As a starting point of the computer information management system, management information systems of endowment insurance in Hebei Province should achieve the following objectives: handling automation of front business. At present, the straight tube units Social Security Information Management of Social Security Administration in Hebei Province mainly use the relevant information gathered by social information administrator of the units. Generally using paper as the media, then the computer professionals of the Social Security Administration re-check and enter the information into the computer which exists in the paper, and it has the character of heavy workload and repeat work. The existing system has been difficult to meet the needs, has been difficult to support more and more onerous insurance business, the construction of the new system is to achieve the handling automation of front business, greatly improving the efficiency of the handling automation of front business, in order to adapt to the social security business development needs; standardize and optimize business processes.

2 Building of the Pension Insurance Management Information System

2.1 Overall Design Ideas

Making pension insurance relationship as the core, the insured individual as the starting point, we should deeply analyze the properties of factors that affect the relationship between the pension insurance, the relationship between attributes, conditions of properties change and conversion, links and the main body of the change and transformation, a thorough analysis of pension insurance relationship and its properties, then according to the organic contact of pension insurance relationship, establish tight association and constraints between data and the business processes, ensure the premise of the complete business, effective restraint and process optimization, optimize the design of business processes, provide flexible combination of increasing and decreasing under various conditions, So that the system has sufficient adaptability.

2.2 System Requirements Analysis and Related Technologies

Through a unified data exchange platform, provide a unified interface solutions, and endowment insurance in Hebei Province Network Management Information System should be taken into account many factors, such as business norms, business development direction, available resources, to overcome the business or technical difficulties, the goal of the system construction, time schedule constraints at all levels use the object, etc. In order to achieve the connection with the external systems of finance, banking, insurance, we should use the method of web services, call interface, and DLL to achieve external providing a unified interface. Based on J2EE technology architecture system, the system using the framework of three-tier architecture based on J2EE, we can extend the core platform second version of the framework. The system can support traditional client by SOPA agreement, use XML and application servers to interact, and also support the direct connection mode of the browser, because of using EJB components, so it supports for the cluster technology. Aim at each type of retirement, also divided into the method of calculation, at present; it can be summarized as follows (Table 1).

Table 1. Calculating table of basic endowment insurance management information

Code	Type	Conditions and notes
1	calculation of pension schemes in Hebei Province	all insured
2	enjoy a one-time treatment of retirement	contribution years less than 15 years is special circumstances of pension calculation in Hebei Province
3	retirement rules of male> 50 and female> 45	male> 50 and female> 45, special trades or early retirement due to illness
4	retirement rules of male<50 and female<45	male<50 and female<45, special trades or early retirement due to illness

3 Architecture Design of Pension Insurance Information Management System

3.1 The Module Design of System Function

Endowment insurance management information system is mainly around the business processing functions of basic information management ,pre-information, data review and statistical queries in endowment insurance management information, builds a set of advanced technology, reliable performance, full-featured system, does the system their work within their respective mandates while ensuring the flexibility , ease and practicality of the workflow, and has a high degree of stability, good scalability and reusability; strict access control, security system security. See figure 1.

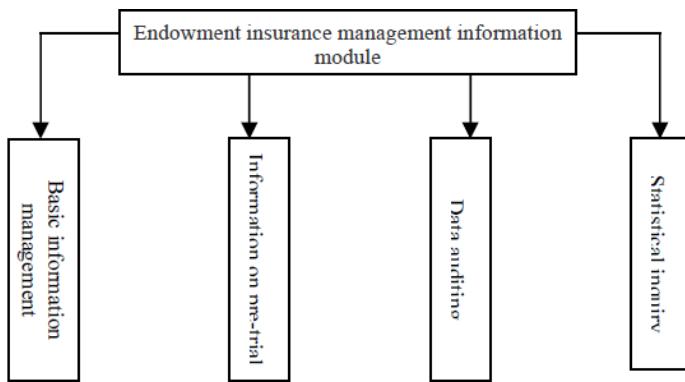


Fig. 1. Endowment insurance management information module

3.2 Technical Principles

During the time of the network system planning, the design should base on the following technical principles: (1) standard: (2) practicality: (3) Reliability: (4) advancement: (5) open: (6) security: (7) scalability: (8) maintainability: (9) ease operating:

3.3 Software Architecture

The software architecture is a structural factor, which is a collection of components, including some form of processing elements, components, and data connectivity components. The processing section is responsible for data processing, the data processing part uses information to connect the components, and the different parts of packet structure are connected by the components. This definition distinguishes centralized processing components, data components and connects components, on the other definitions and methods remained essentially unchanged. Based on component architecture, the relationship between components, the relationship of

environment between components and the basic organizational structure content of the specific system, it is the principle in the design and principle of evolution. Client / Server distributed computing architecture, the application shares the processing between client and server, a request is usually processed by a relational database, the PC machine is to receive the processing and display data to business logic, and the system supports modular development. Client/Server due to its flexible architecture is widely used. Development of the Internet has had a profound impact in terms of the traditional application development. Internet and network software applications require a more open and flexible architecture. A new and more viable three-tier / multi-tier computing architecture has been widely used, it is composed by the server layer, client layer and data layer, the portability of the three-tier structure is more advantageous, and it can cross different types of platform, while allows the user to request load balancing between multiple servers. The three-tier structure is easier to achieve security, because the application has been isolated client. The application server is a three-tier / multi-tier architecture, its components are the location of the middle layer. See figure 2.

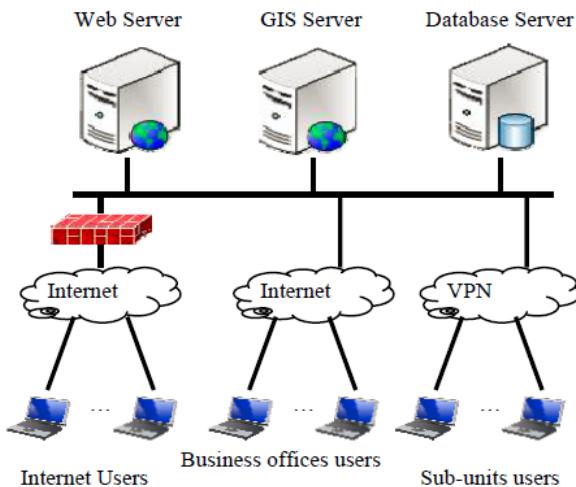


Fig. 2. The C / S model schematic based on the three-tier structure

3.4 Software Platform and System Security

Software platform: As shown, it uses the third-party mainstream products as the middle layer of the three-level network application architecture. In transparent state of applications system, it achieve vertical expansion of cross-platform .Especially in the case of e-commerce type applications are difficult to fully expected to load on the system, greatly improving the flexibility of the system. It combines with other excellent software platform, and fully meets the requirements of the enterprise-class applications.

Main technical: In the preparation of the application logic, it mainly uses Java language standard of HTML, Java Servlet, JavaBeans, JSP, EJB, XML, Struts, hibernate and spring, also implicated in technology or protocol of JDBC, ECI, EPI, RMI, IIOP and WAP.

System Security: The security design of multiple firewalls divides the entire network infrastructure into physical and logical, in order to make configuration system to reach the C2 level of security.

For the details see figure 3.

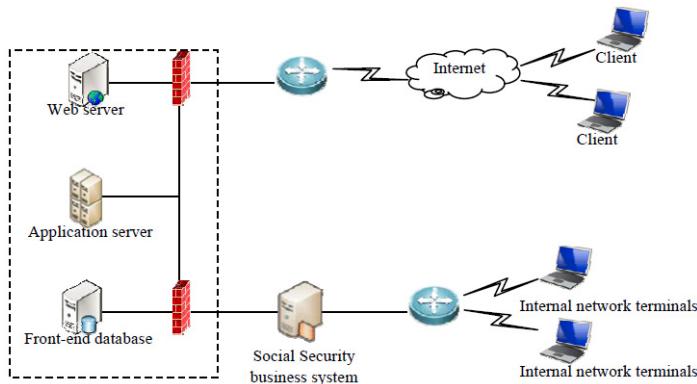


Fig. 3. The security design of multiple firewalls

4 Design of Database Platform

In accordance with the unified planning, it uses the centralized management model of the Provincial Social Security Bureau for the center, and the following set up the direct control of the units do not establish a data center. Centralized management model is help to save investment, unified policy, regulate business, facilitate statistical analysis, analyze decision, and achieve the fund monitoring. Under the centralized management mode, the system design takes full account of information sharing to staff across-unit transfer.

4.1 Design of Basic Service Platform

Business application supporting platform includes the underlying application framework and foundation management services. The application framework is developed based on J2EE, Use of the core platform, and uses the second version of the B / S / S structure. The development of basic management services platform include unified user management, business step by step authorization, generic parameter management, general inquiries and other functions.

4.2 The Scope and Content of a Unified User Management Labor and Social Security Information System Interconnection are Growing

A complete set of social security information systems will inevitably need the user management mechanism and management platform of comprehensive, unified, based on the combination of administrative and business applications. Ensuring that it can easily expand the user's content and permissions also be seamlessly with the business system docking, and achieving business expansion and function expansion of application. To this end, we propose the management mechanism of fitting actual situation of the social security industry, covering all applications, unified management, unified interface to manage authentication and unified user permissions control management. Unified user management includes: business management, user-category management, user management, user group management, user login management and so on. See figure 4.

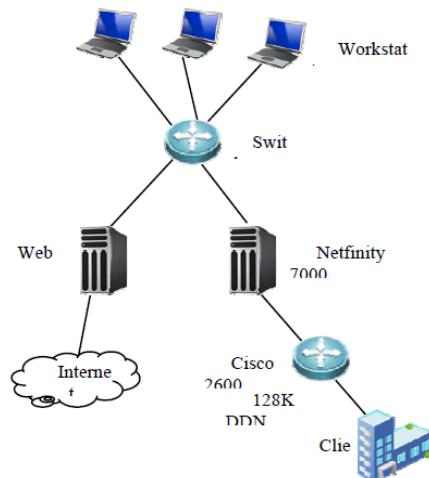


Fig. 4. Network management information system in straight old-age insurance

4.3 The System Should Provide a Unified Common Query Capability, According to Users' Changes in Demand, Free to Set Their Own Search Requirements and the Final Data Performance

Unified General inquiries should be scattered in every corner of the query to focus on finishing, so that they can appear as a whole. Using a common query tool to solve the problem of the large amount of data and demand variable query, with a simple method to quickly and easily generate or modify a query, it can greatly enhance the value of operation of the system and reduce system maintenance costs. Unified query with the following features: back-office services simplicity: the back-office services through the common query data table structure to achieve configuration; comprehensiveness of the data retrieval: the query tool can query simple message of

the general trading, and make simple statistics to some complex queries, such as sum, average, maximum, and minimum; easily browse of data performance: general Query tool provides powerful browsing capabilities; easy output of query results: it provides a variety of output processing on the query results, you can generate a print file or a simple text format ,also can be printed directly or with relevant view tools for reading, and provides advanced features of the output to EXECL files; the ease development process: it through a simple text configuration to completed development of a query ; database independence: general query tool has nothing to do with the database, the retrieved data source connect with platform, and retrieving through the platform provided by the data source; statistical analysis of the retrieve the results :it can be divided into A: seriously affect system operation; B: affect system operation; C: does not affect the operation but must be modified; D: the recommendations.

4.4 Design of System Application Software and Hardware Platform

The system uses an advanced system architecture based on JAVA technology, built on WebLogic8. X Server, and Oracle 109 database, the development tools are Myeclipse5.5 and Jdk1. 4 lib, applying the most popular SSH framework combination, that are the Struts, Hibernate, and the spring combination. WAN platform is the VPN virtual private network.

4.5 Design of System Security

According to network structure and network firewall performance requirements, we select hybrid technology and firewall products with two or three port, use the structure of dual-host, firewall, routing and encryption algorithms to enhance system security.

5 Conclusions

The system design is reasonable, stable, easily to operate, and share data with other insurance systems. In this paper, we design and implement endowment insurance management information of Hebei Province. The system has passed the acceptance, applies in each provincial unit, and realizes standard, automation, and scientific management of endowment insurance management modernization. The system provides fast, reliable basis for the leadership on macro decision-making, greatly reduces work burden of the basic units and improves the level of information management services. Although the design and realization of the endowment insurance management information systems in Hebei Province is success, however, the entire system is more complex, technology and knowledge involved in are more, workload is heavy, the work done must be some things consider incomplete and imperfect, so there are still some work that require further to study.

Acknowledgments. The research was supported by 2012 Social Science Fund Projects of Hebei Province (HB12SH029).

References

1. Galbraith, J.: Organization Design: An Information Processing View. *Interfaces* 4, 28–36 (1974)
2. Jin, Y., Levitt, R.E.: The Virtual Design Team: A Computational Model of Project Organizations. *Journal of Computational and Mathematical Organization Theory* 2(3), 171–195 (1996)
3. Levitt, R.E., Cohen, G.P., Kunz, J.C., Nass, C.I., Christiansen, T., Jin, Y.: The ‘Virtual Design Team’: Simulating How Organization Structure and Information Processing Tools Affect Team Performance. In: Carley, K.M., Prietula, M.J. (eds.) *Computational Organization Theory*. Lawrence Erlbaum Associates, Hillsdale (1994)
4. Wegner, D.M.: Transactive Memory: A Contemporary Analysis of the Group Mind. In: Mullen, B., Goethals, G.R. (eds.) *Theories of Group Behavior*, pp. 185–208. Springer, New York (1987)
5. Atkins, D.E., et al.: Revolutionizing Science and Engineering Through Cyber-Infrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure. Washington, D.C. (January 2003),
<http://www.cise.nsf.gov/sci/reports/toc.cfm>
6. Besek, J.M.: Copyright issues Relevant to the Creation of a Digital Archive: A Preliminary Assessment. Washington, DC: Council on Library and Information Resources and Library of Congress, National Digital Information Infrastructure and Preservation Program (January 2003), <http://www.clir.org/pubs/reports/pub112/pub112.pdf>
7. Bimber, B., et al.: The Future of Technology and Learning in the University. UCSB Center for Information Technology and Society (March 2002),
http://www.cits.ucsbs.edu/site/research/cits_wp_32002_1.pdf
8. Bollen, J., et al.: Usage Analysis for the Identification of Research Trends in Digital Libraries. *D-Lib Magazine* 9(5) (May 2003),
<http://www.dlib.org/dlib/may03/bollen/05bollen.html>

Library Microblogging Based on Sina Microblogging Platform

Yang Shuangqi, Xu Jing, and Shao Shengchun

Library, Hebei United University, Tangshan 063000, Hebei, China

Abstract. Microblogging is the Web2.0 technology with the most important impact in the recent years, with applications in many domains. At present, many libraries and information science experts in China launched microbloggings. Our paper takes Sina microblogging as the survey platform, takes certified library microbloggings as the objects. Analyses the results of the survey carried out on 224 Chinese library microbloggings. Brings forward the problems faced by library microbloggings and gives improving methods.

Keywords: Library, Microblogging, Internet survey, Information service.

1 Introduction

Microblogging is a Web2.0 technology and a relationship-based information sharing and accessing platform, which allows the users to publish online brief text updates, usually almost 140 characters. Microblogging enables a real-time interaction between users, by the using of mobile devices anywhere, anytime, by anyone and anything. Usually, microblogs are readerbased (also called contact-based or pull-based) social networks. The earliest and best known microblogging services are Twitter of the United States, the product already has 75 million registered users in the world. In August 2009, China's largest portal Sina launched Sina microblogging private beta, becoming the first microblogging service portal site, microblogging officially entered the mainstream vision of Chinese Internet.

According to the China Internet Network Information Center (CNNIC) report, up to the end of December 2011, the number of microblogging users in China reached 250 million, with an increase of 296.0% compared to the end of 2010, the use rate of internet users was 48.7%. Microblogging has developed into an important Internet application used by nearly half of Chinese Internet users. Microblogging with ease, simplicity, fast, instant, easy access and interactive features, can provide a relaxed exchange environment, for libraries to provide information services to users easily. It also allows users to find a platform to express free expression of personal experience, reading experiences, information needs, as well as the proposed for the library. Microblogging has become a choice of many libraries to publicize and carry out information services.

2 Survey Method and Content

The survey was carried out in April, 2012. Select Sina microblogging which has the largest number of domestic users as the research platform, use "library" as the searching key words. Take the certified microblogging users as research object. Investigate the circumstance of how libraries open microblogging to provide services. Specific investigations include the library microblogging usage and services. Bring forward the problems library microblogging have and gives improved measures based on the analysis situation of the status in authenticated microbloggings.

3 Results and Analysis

Up to April 2, 2012, a total of more than 500 microbloggings use the name of "library", among them there are 482 Users being certified by Sina. Through further filter, we exclude profit organizations, periodicals newspaper websites and entities that are inconsistent with library institutions. At last, we got 224 libraries microblogging remain. Except for 27 libraries personal microbloggings, the applicants are mostly library curators or industry experts, the rest are all institution microbloggings.

3.1 Basic Status of Library Microblogging Type

The number of university libraries and public libraries was both very large in all the institution microblogging proposers. We also found many social welfare library and virtual digital library; the specific data is shown in Table 1.

Table 1. The types of Sina certified library microbloggings. (N =224)

Library type	University libraries	Public libraries	Public libraries	welfare libraries	Professional libraries	Digital libraries
Library number	79	82	20	7	7	
Proportion (%)	35.3	36.6	8.9	3.1	3.1	

3.2 Library Microblogging Locality

Among the 224 library-related certification microbloggings which being researched, the applicant region covers nearly all provinces. Beijing has the largest number of Certified Library microbloggings of 45, followed by Shanghai and Guangdong, respectively, 29 and 27. Qinghai, Tibet, Jiangxi and Jilin haven't opened a certified library microblogging. In east China there is 85 certified microbloggings, occupying 37.9%. The specific data was shown in Table 2.

Table 2. Locality of Sina certified library microbloggings. (N = 224)

Library type	East China	South China	Central China	North China	Northwest	Southwest	Northeast
Library number	85	19	16	61	18	15	10
Proportion (%)	37.9	8.5	7.1	27.2	8.0	6.7	4.5

3.3 Basic Situation on the Number of Microbloagging Fans

The number of fans is a important indicator in measuring user's attention to the Library microblogging. First and foremost, a microblogging has the potential to foster a sense of community within and beyond the walls of the library. Library microblogging fans can be divided into several categories, students, the information science experts, other agencies of the school and other peer units. The survey results show that the library microblogging is mostly being well concerned .there are 37 microblogging whose fans have been more than 5000, this indicate that the Library microblogging were generally well received by users. The specific data is shown in Table 3.

Table 3. Sina certification library microblogging fans number. (N = 224)

The number of library fans	1 ~	200 ~	500 ~	1000 ~	2000 ~	5000 ~
Library number	23	40	37	39	48	37
Proportion (%)	10.3	13.8	16.5	17.5	21.4	16.5

3.4 Basic Status of Number of Microblogs Being Published on Microblogging

The content of microblogging messages goes about activities of one library, opinions and status, varying from sharing activities and accomplishments with users, relatives and co-workers; sharing news and opinions with interested readers; and seeking knowledge and expertise in other public messages. The structuring of discourse in communication tools represents the possible relationships among messages. The survey result shows that until April 2, 2012, among the 224 investigation object ,All of the users published at least 3 microblogs, the highest one wrote 7483, more than half of the library microblogging have published more than 200 microblogs, the specific data is shown in Table 4.

Table 4. The number of messages being published on Sina certified library microbloggings. (N = 224)

number of messages	3 ~	50 ~	200 ~	500 ~	1000 ~
Library number	35	48	56	41	44
Proportion (%)	15.6	21.4	25	18.3	19.6

3.5 Library Microblogging Service Content Analysis

The services provided by libraries on microblogging platform can mainly be divided into several kinds, such as news, resources introduction, push service, reference, education and training, subject service and promotions. It's easy for users to accept microblogging, serving as a quick information publishing platform. In its wide range of services, News services are provided by almost all library microboggings. Resources introducing and reference consulting services is also commonly carried out, some libraries directly refer the consultation to be the microblogging service positioning. In additional, many libraries microblogging provides push delivery service, including bibliographic recommendation, trial database, and recommendation of the digital media resources and so on. While the development of Subject Services is not wide enough, only a few Library microboggings put forward this kind of service. A variety of services carried out by the specific circumstances in Table 5:

Table 5. Sina Certification Library microblogging service content analysis.(N = 224)

Service type	News	Resources introduced	Push service	Reference	Education and training	Subject Service
Library number	221	126	72	95	56	23
Proportion (%)	98.7	56.3	32.1	42.4	25.0	10.3

4 Problems and Countermeasures

Microblogging, the act of broadcasting short, real-time messageshas been embraced world-wide and is the most representative case of reader-based social networks. Several studies highlight microblogs as an increasingly popular and socially acceptable means of information exchange. As described in the paper, though there are extreme advantages in using a microblogging, and the development speed is very fast, there are yet many practical problems which have to be solved, the service model remains to be further. These are the main problems and related measures in libraries using microblogging.

4.1 Make Microblogging Objects and Functions Clear

Many libraries use the microblogging for service with its functional orientation was unknown, this can't do good to the development of services and for more people to use it. Before library registers apply the microblogging, there must be a development plan, which can help the managers to determine the target groups of the microblogging and the proposed goal. The content and structure of microblog, writing requirements, as well as the microblogging working staff and manager should have a clear position. Libraries should take great efforts to ensure the microblogging being continuous renewed and have many new thoughts to attract more users.

4.2 Certify Microblogging and Standardly Set the Microblogging Name and Label

For an institutional microblogging, the post content must be true and authoritative. As we have stated, the general assumption in works such as these is that social presence serves as the basis for building successful communities. Library official microblogging should be certified as soon as they can. Authentication in the application can help users to find the library they like easily and assure the authenticity of microblog contents. Microblogging name should be complete and specific; the introduction should include brief description of the Library's web site, blog link address, telephone number, etc. It will be the best if the library put microblogging content and service instructions in the microblogging tags. In this way libraries can facilitate the users to find the microblogging, give efficient attention and participate in the interaction in the long run.

4.3 Deepen Library Microblogging Service Content

From the Library microblogging survey data, we can see that, although libraries provide many kind of service and have got some effects, the consulting service, push service as well as the subject service still needs to be furthered. Libraries should fully understand the advantages of microblogging as real-time interaction and initiation in information pushing. Microblogging can be seen as the ideal way to promote reference and push service in the Web2.0 era. Take reference consulting service as the main service content provided on microblogging. Encourage the subject librarian establishing different theme of microblogging depending on the variety range of subjects, and eventually make a formation of library science microblogging system, provide readers with a deeper subject service.

4.4 Pay Attention to the Construction and Maintenance, Strengthen the Propaganda Power

Currently, library microblogging is showing a large scale of development, nevertheless, some microbloggings were in poor quality situation, the libraries have an uneven microblog writing frequency. To change this situation, requires emphasis on microblogging construction and maintenance of libraries in terms of ideology, microblogging Applicants can use the full-time personnel to management microblogging, set a minimum update frequency, the shortest reply time limit. Enrich microblogging content to attract more readers concerned about the microblogging. Strengthen the propaganda power, organize interactive activities on a regular period of time, and improve the cohesion of the microblogging. Use tools such as microblogging group to tap the potential library users, and keep on service innovation, attract library users to follow the Library in a long-term, and improve the libraries' Popularity in the user group.

5 Conclusion

Microblogging is an information distribution center, it creates a network, cluster, integrated information dissemination among users, on which information is highly

closely together. It has been less than two years since the launching of the first library microblogging, however, library microblogging has made a great progress in terms of the service content and the effect. Microblog is brief and can be received and viewed through the Internet, mobile phones and a variety of ways. Microblogging, proved to be an effective tool for professional development and for collaboration with users. Furthermore, as a social networking platform, microblogging provides valuable interactions between libraries and their users, acting as a social factor in a management system. The information service provided on microblogging is far convenient than the service offered in traditional methods. Libraries should make effective use of the features and advantages of microblogging in information dissemination, knowledge sharing and user integrating. Realize the value of microblogging service, provide users with a deep-seated, high-quality information services.

References

1. China Internet Network Information Center (CNNIC). Statistical report on the development of Internet in China.
http://www.cnnic.cn/research/bgxz/tjbg/201201/t20120116_23668.html
2. Chen, G.M.: Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*, 755–762 (2011)
3. Xin, G.: Microblogging in the library literature publicity vision. *Modern Intelligence*, 54–56 (2012)
4. Huang, S.: Library microblogging use characteristics and development strategies. *Journal of Academic Libraries*, 78–83 (2012)
5. Lin, C.: Domestic Library microblogging application status. *Library Science* 5, 30–33 (2011)
6. Chao, L.: Domestic Library opened status of the microblogging service. *Library Journal*, 24–28 (2012)
7. Jing, L.: Analysis of university library to authenticate the user microblogging survey - Sina microblogging platform. *Library Science* 01, 90–95 (2012)
8. Ping, J.: Microblogging explore in the University Library. *Science and Technology* 27, 600–618 (2011)
9. Li, J.: Foreign libraries micro-blog building and Its Implications. *Library and Information* 01, 70–73 (2011)
10. Li, J.: Microblogging in the University Library Information Service. *University Library* 04, 44–46 (2011)

XML-Based Web Data Pattern Discovery and Extraction

Rui Jia^{1,2}, Shicheng Xu^{1,2}, and Chengbao Peng^{1,2}

¹ Neusoft Corporation, Xinxiu Road, Shenyang, P.R. China

² Northeastern University, Wenhua Road, Shenyang, P.R. China

{jia.r,xushicheng,pengcb}@neusoft.com

Abstract. This paper presents an XML-based web data extraction method. This method translates web page into XML document, analyze XML document by using XPath/XSLT, discover web page data pattern and similarity by using XML clustering algorithm, construct XPath-based data extraction rule template. This method improves the robustness and versatility of web data extraction system. Experiment result shows that the data extraction method has high precision and is adaptive to web pages in different sites and with different structures.

Keywords: Web data extraction, XML Clustering, Pattern Discovery.

1 Introduction

With the development of worldwide Internet, get useful information from the Internet becomes difficult. The Web data are vast amounts, heterogeneous, dynamic changes. And most Web data are in HTML form, they lack data description and clear semantic information. These shortcomings make Web data extraction different from traditional information extraction. Web data extraction becomes an important research field of information extraction.

There already have some researches on Web data extraction and these extraction methods and models have solved some problems in Web data extraction, improve the development of information extraction technology. But there still have some limitations or shortcomings in these extraction methods. Therefore, it is necessary to do further research on Web data extraction technology to enhance its performance and efficiency.

2 Web Data Extraction Overview

Information Extraction refers to the automatic extraction of structured information such as entities, relationships between entities, and attributes describing entities from unstructured sources. Information extraction can help users to find and analyze information easily. Users can get useful information effectively and do further data extraction or text analysis.

Web data extraction arises with the development of network technology and the expansion of web messages. Web data extraction develops through manually, semi-automatic to automated operation. Web data extraction identifies data from unstructured or semi-structured information data in Web pages and translates data into structure and semantics format.

Wrapper is an important component of Web data extraction system [1]. Wrapper is a software process that uses pre-defined extraction rules to extract data from collected Web pages, and converts data to specific information format. Wrapper can be considered as a program or rules to understand source information. It translates information data to specific structure format, such as XML or relational tables. Wrapper is specific for a particular given site. Because Wrapper is closely related with the web page structure, so the challenge for Wrapper is to identify useful data from many unrelated pages [12].

3 XML in Web Information Extraction Application

XML is a simplified subset of SGML. XML does not care about data presentation, it only cares data description. An XML document is structural, which makes it fit for data description, exchange and interoperability. XML has been developed quickly and many XML related tools have been developed. These advantages make XML based applications are more efficient, robust, and easier to develop or maintain.

XML data model are suitable for semi-structured data because it can combine data from different sources easily. XML-based web data extraction method provides a solution for web semi-structured data model and data extraction problems. Use XML in web data extraction can achieve simple, efficient, robust goals.

In the data extraction process, XML works as data exchange, it is easy to parse and process. The main task of data extraction is to find out Wrapper's mapping rules. XSLT can be used as Wrapper to extract data from XML documents. XSLT can map one XML document to another XML document with different format.

XPath is an expression language to locate address in XML document. It provides a common syntax and semantics between XSL transformation and XPointer. XPath works on abstract logical structure of XML documents. XPath provides address location, string manipulation, number and Boolean functions, also design a matching natural subset which is used for matching purpose.

4 XML Clustering

XML text clustering divides XML document sets into different groups, which makes documents in same group are similarity as much as possible. Because XML is a structural document, its semantic information can be described by document structure. XML clustering can be done by analyzing and comparing XML document structure.

For XML clustering [8], we have three methods to calculate the document similarity: elements comparison, edge sets comparison and distance comparison method.

Elements comparison

Calculate the ratio of same element number between two XML documents. This ratio reflects the similarity of these two XML documents. Because high level elements (which are close to the root element) reflect document structural information better than low level elements (elements which are away from the root element), so we must put element level into account when calculating document similarity.

Edge sets comparison

Calculate the ratio of same edge number between two XML document trees. This ratio reflects the similarity of these two XML documents. Obviously, more same edges mean two documents are more similarity.

Distance comparison

Calculate the edit distance between two XML document trees to measure similarity. This method define two trees' distance is the cost which uses edit operation to translate one tree to another tree. Obviously, XML document distance and similarity are inverse relationship. Three main operations are: insert, delete, and update.

5 Web Data Extraction System

5.1 Purpose of Web Data Extraction System

Research existing Web data extraction techniques and analyze their advantages and disadvantages. Manually Web data extraction system has higher accuracy, but requires more labor and extraction pattern analysis ability. Also wrapper maintenance is a time-consuming task. Semi-automatic [10] data extraction system requires lots of samples to do pre-study, it needs less operator interaction. Automatic data extraction system minimizes manual involvement and constructs extraction patterns by analyzing web page similarity.

Purpose of Web data extraction system is:

Construct robust and flexible wrapper to reduce operator interaction, improve extraction efficiency and automation capability.

Wrapper construction should be simple and easy to be operated to adapt the changes of web data and forms.

5.2 Web Data Extraction System Design

In this paper, we propose a XML-based web information extraction system which combines automatic information extraction technology [5]. This system uses XSLT language to describe extraction rule, use XPath expression to locate data address, generate extraction rules to construct wrapper. Use location-based and content-based XPath expression to improve extraction performance and robustness according to different situations.

Extraction system uses XML document clustering algorithm to analyze web pages' similarity and constructs new extraction rule template automatically by reusing existing extraction rules, which improves the scalability and adaptability of whole extraction system [13].

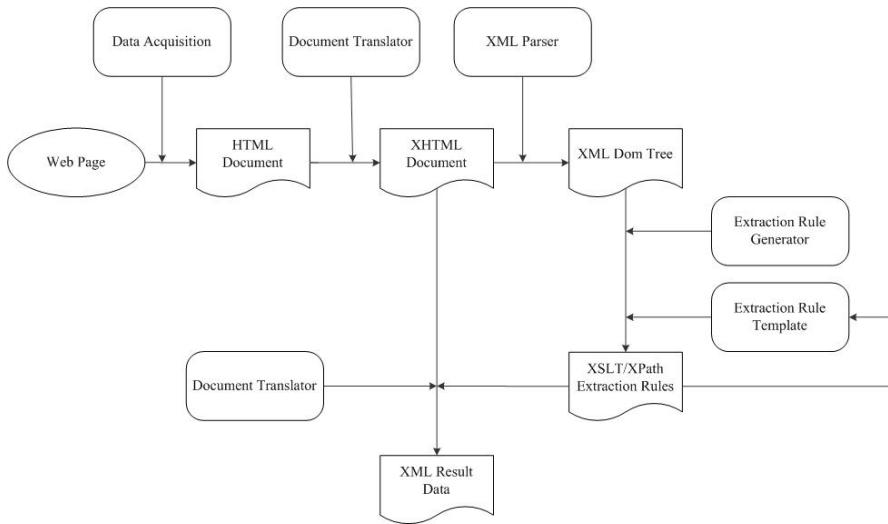


Fig. 1. This figure shows architecture and components of web data extraction system

Web data extraction system has following components:

Data Acquisition: Obtain Internet web pages according to user's needs, clean up page to remove useless information.

Page Parser: Translate non-structure HTML document to XHTML document, which are well-structure and in accordance with XML standard. Use XML parser to translate the XHTML document to XML DOM tree [3]. In Web data extraction system, python module chilkat [9] is used to translate HTML to well-structure XHTML, and python module xml.dom.minidom is used to construct DOM tree.

Extraction Rule Generator: Mark useful information data from user point of view. Use XML document clustering algorithm to do similarity comparison, generate extraction rules automatically by reusing existing extract rules. Use rule learning process and position location optimization method to improve the web data extraction rules. Get XPath expressions for data nodes which will be extracted, construct XSLT extraction document. Store new extraction rules into the extraction rule template library.

Data Extractor: Search existing extraction rules to see if there are any suitable templates when user requires extracting data. If there is no suitable rule, re-learn page structure, construct and store new extraction rules. The extraction results will be presented in XML format.

5.3 XSLT Extraction Rules Construction

Web data extraction system using XSLT as description language, and node position location is based on XPath expression in XSLT [11]. Extraction rules should be more robust because the data or structures of web page always change. So, extraction rule generator must put these problems into consideration to avoid extraction rules failure.

In common situation, XPath expression is always based on tree path, such as /html[1]/body[1]/table[2]/tr[5]. This kind of expression is the most intuitive, but it's suitable for HTML documents because the basic HTML tags are based on presentation, such as table. Different data often use the same tags, such as tr, td. Then, XPath expression is likely to be changed when add or delete data and structure in HTML document. For example, the last XPath express will be useless if we add another line in the table, the extraction system will fail if the wrong extraction rules are still being used. These kinds of problems bring risk and lower flexibility to web data extraction system.

Therefore we involve two methods: content based location method and relative position based location method to enhance web data extraction system, improve extraction rules' robustness [2] and effectiveness [6].

Content based location method uses unchanged page content or user needed information to locate data address. For example, operator wants to extract which line contains 'Neusoft' from table, then use XPath expression //tr[contains(normalize-space(), 'Neusoft')]. This XPath expression makes sure the accuracy of extraction rules even if web page structure has been changed.

Content based location method has advantages: high validity, easy to understand, automatic adaptability. Content based location method does not depend on web page structure, so extraction rules construction won't be affected by page changes. Extraction rule constructor doesn't need to understand the complex structure of whole page. Relatively speaking, the shortcoming of content based location method is that the performance is not good as the direct path location method, since it has to search and compare key word.

Besides content based location method, web data extraction system also introduces relative position based location method. Relative position based location method is based on Anchor-Hop model. Anchor is a reference node, usually contains the most recent common ancestor for all information need to be extracted. Then use relative path to locate the information needed to be extracted, which is hop. Because the relative path is very short, so it is not sensitive. By combining content based location method and relative position based location method, performance of web data extraction system can be improved.

The following XLST extraction rule document presents how to extract data for Neusoft corporate sector information. First, system locates 'Neusoft' anchor node by using content based location method, and then extract the department name, description and product information by using relative position based location method which is implemented by XPath following functionality.

Example of extraction rule document

```
<?xml version="1.0" encoding="gb2312"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output encoding="gb2312" indent="yes"/>
<xsl:template match="/text() [contains(., 'Neusoft')] ">
<department-1>
```

```

<name>
<xsl:value-of select="following::text() [1]" />
</name>
<description>
<xsl:value-of select="following::text() [2]" />
</description>
<products>
<xsl:value-of select="following::text() [3]" />
</products>
</department-1>
<department-2>
<name>
<xsl:value-of select="following::text() [4]" />
</name>
<description>
<xsl:value-of select="following::text() [5]" />
</description>
<products>
<xsl:value-of select="following::text() [6]" />
</products>
</department-2>
</xsl:template>

```

5.4 Web Data Extraction Clustering Formula

5.4.1 Elements Comparison Formula

Element (T) = { $E_1, \dots, E_i, \dots, E_n$ } presents all elements in XML document tree. Here are two XML document trees T_1 and T_2 , Common (T_1, T_2) = { $C_1, \dots, C_i, \dots, C_m$ } presents all same elements between T_1 and T_2 , each element can be presented as $C_i = E_{1j} = E_{2k}$, LevelC (C_i) = Max (Level (E_{1j}), Level (E_{2k})). Common' (T) = { $C'_1, \dots, C'_i, \dots, C'_n$ } presents different elements between T_1 and T_2 . The formula which calculates the element similarity between T_1 and T_2 is:

$$Element(T_1, T_2) = \frac{\sum_{i=1}^m LevelC(C_i)}{\sum_{i=1}^n Level(C'_i) + \sum_{i=1}^m LevelC(C_i)} \quad (1)$$

5.4.2 Edge Sets Comparison Formula

T_1 and T_2 present two XML document trees, $Edge(T_1)$ and $Edge(T_2)$ sets present the edges in these two trees. $Edge(T_1) \cap Edge(T_2)$ contains the same edges between T_1 and T_2 , $|Edge(T_1) \cap Edge(T_2)|$ is the number of same edges. $Edge(T_1) \cup Edge(T_2)$

contains all the edges between T_1 and T_2 , $|Edge(T_1) \cup Edge(T_2)|$ is the number of all edges. The formula which calculates the edge similarity between T_1 and T_2 is:

$$Edge(T_1, T_2) = \frac{|Edge(T_1) \cap Edge(T_2)|}{|Edge(T_1) \cup Edge(T_2)|} \quad (2)$$

5.4.3 Distance Comparison Formula

Calculate the edit distance between two XML document trees [7] to measure similarity. Each tree edit operation has cost, Distance (e) presents the cost value of one edit operation, and this value should be a positive number. T_1 and T_2 present two XML document trees, and $Edit = \{E_1, \dots, E_i, \dots, E_n\}$ presents the edit sequences which translate T_1 to T_2 . So, the edit distance is the sum of each edit operation's cost. The formula is:

$$Dis tan ce(Edit) = \sum_{i=1}^n Dis tan ce(E_i) \quad (3)$$

There may be many ways to translate T_1 to T_2 , the minimal translation sequence is defined to be the edit distance between T_1 and T_2 . The formula which calculates the edit distance similarity between T_1 and T_2 is:

$$Dis tan ce(T_1, T_2) = Min(\sum_{i=1}^n Dis tan ce(E_i)) \quad (4)$$

5.4.4 Extraction Clustering Formula

Web data extraction system combines three XML clustering formulas to analyze page similarity [4]. Because each clustering formula has advantages, combine all of them can improve accuracy. So, clustering formula for extraction system is:

$$S(T_1, T_2) = \alpha * Element(T_1, T_2) + \beta * Edge(T_1, T_2) + \gamma * Distance(T_1, T_2) \quad (5)$$

In this formula, $\alpha + \beta + \gamma = 1$. User can change these parameter values or extraction system will change them according to page characteristic.

5.4.5 Test Data

Use web pages from Neusoft, Baidu and Google as test sets. Analyze and compare page similarities to verify the accuracy of web data extraction system.

Experiment result shows that the extraction system has high precision and is adaptive to web pages in different sites and with different structures. The pages from www.neusoft.com site have higher similarity, which are 80% or more, then web data extraction system can reuse extraction rules template for these pages. On the other side, the similarities between Neusoft and Baidu/Google are low, which are less than 30%. These indicate the website structures are different, existing extraction rule templates can't be used.

6 Conclusion

This paper proposes a XML-based web data extraction system. This extraction system uses XML as data exchange method, also use XML to describe semi-structured data model. Extraction system uses XSLT and XPath as the extraction rule description language, and combines XML clustering analysis to construct data extraction rules. The advantage of this system is that data extraction rules generation process needs less manual interaction, improve automation and can adapt different web page structures. In the near future, we will do further work to improve extraction and automation, build more robust extraction rules and continue improving the performance and accuracy of XML clustering algorithms.

References

1. Ion, M., Steven, M., Craig, K.: A Hierarchical Wrapper Induction for Semi structured Information Sources. *Autonomous Agents and Multi-Agent Systems* 4(1-2) (1999)
2. Jussi, M., Jared, J.: Robust Web Data Extraction with XML Path Expressions. IBM Research Report (2002)
3. Shoubiao, T., Chao, X., Yuan, J.: Web Data Extraction System Based on Label Library. In: *Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 7 (2009)
4. Wei, L., Xiaofeng, M., Weiyi, M.: ViDE: A Vision-Based Approach for Deep Web Data Extraction. *IEEE Transactions on Knowledge and Data Engineering* 22(3) (2010)
5. Arvind, A., Hector, G.-M.: Extracting structured data from web pages. In: *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data* (2003)
6. Jussi, M.: Effective Web Data Extraction with Standard XML Technologies. In: *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong (2001)
7. Yuan, W., David, J., De Witt, J., YiCai.: An Effective Change Detection Algorithm for XML Documents. In: *Proceedings of the 19th International Conference on Data Engineering* (2003)
8. Pan, Y.: Research on XML Documents Cluster. *Information Journal* 25(2) (2006)
9. Python chilkat, <http://www.chilkatsoft.com/python.asp>
10. Xu, J., Hou, K.: Autonomous Extract Information from Web Pages. *Computer Engineering and Applications* 14 (2005)
11. Ling, L., Calton, P., Wei, H.: XWRAP: An XML Enabled Wrapper Construction System for Web Information Sources. In: *Proceedings of the 16th International Conference on Data Engineering*, San Diego, California (2000)
12. Zhao, S., Chen, S., Deng, S.: Design and Implementation of Web Data Integration Wrapper Based on Rule Tree. *Computer Technology and Development* 16(6) (2006)
13. Qiang, H., Rui, C., Yanwei, P., Lei, Z.: From One Tree to a Forest: a Unified Solution for Structured Web Data Extraction. In: *34th International ACM SIGIR*, pp. 775–784 (2011)

Application of Decision-Making Model Based on Structure Entropy in IT-Outsourcing Supplier Selection

Li Shi^{1,2} and Hong-bin Dong¹

¹ International School of Software, Wuhan University, Wuhan, China

² School of Information Management, Hubei University of Economics, Wuhan, China
shily0118@163.com

Abstract. A new method based on structure entropy model was presented for selecting the best outsourcing suppliers of IT Corporation. First, the indexes system of selection is proposed. Then, synthesis weights are obtained by using structure entropy method. It belongs to the subjective assignment and objective assignment method based on the entropy theory. Finally, partial large Cauchy distribution function and the logarithmic function are used to be the membership function, each supplier's evaluation results is calculated by fuzzy mathematic method. The best IT-outsourcing supplier is selected according to the sort of evaluation values. An example is given to prove the validity of the method.

Keywords: IT outsourcing, supplier selection, structure entropy, membership function.

1 Introduction

The supplier evaluation and selection of research first began in Dickson [1-2]. Through his survey in 1966, he summed up the 23 supplier selection criteria and weights and their sequencing. The current IT outsourcing supplier selection methods are mainly AHP method [3], fuzzy comprehensive evaluation method [4], nested Logit model [5], etc., and similar partnership also has been discussed by many papers, the use of AHP method [6], AHP-TOPSIS method [7], linear programming method [8-9]. As the IT outsourcing supplier selection contains a lot of uncertainty and ambiguity, which both objective reasons (evaluation of the ambiguity and difficult to quantify, etc.), and decision makers have their own subjective reasons (personality, preferences, values, awareness of the problem, etc.), so the combination of qualitative and quantitative methods, using structural entropy model to select the best IT-outsourcing supplier, which is a new method for the selection of IT outsourcing suppliers.

2 Index System of IT Outsourcing Supplier Selection

In order to make a comprehensive IT outsourcing supplier selection, a comprehensive and scientific evaluation system should be established. On the basis of analysis the

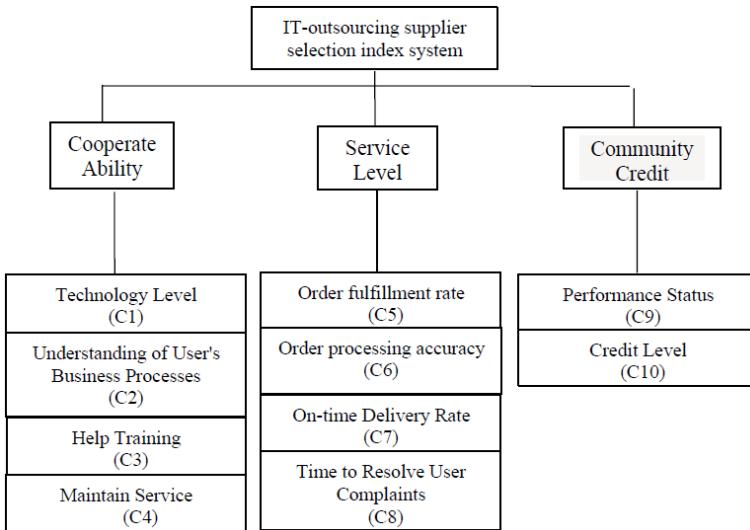


Fig. 1. IT-outsourcing supplier selection index system

relevant references [2-4], a specific IT outsourcing supplier selection system is shown in figure 1.

3 Decision-Making Model Based on Structure Entropy Method

The main idea of this model is: Synthesis weights are obtained by the method of structure entropy. According to the actual need, partial large Cauchy distribution function and the logarithmic function are used to be the membership function, each supplier's evaluation results is calculated by fuzzy mathematic method, and through the sort of evaluation, the best IT-outsourcing supplier is selected.

3.1 Structure Entropy Method [10] to Determine Weights

Step 1: Collecting experts' advices by DELPHI method, forming a sort expert opinion, that is "a typical sort".

Step 2: "Blind-degree" analysis.

Suppose there are k experts, and obtain k tables. Each table corresponds to an index set, denoted as $U = (u_1, u_2, \dots, u_n)$. Index set corresponding to the "typical sort" array, denoted as $(a_{i1}, a_{i2}, \dots, a_{in})$. Sorted matrix is obtained by the k index tables, denoted by $A = (a_{ij})_{k \times n}$, $i = 1, 2, \dots, k$; $j = 1, 2, \dots, n$, which a_{ij} shows the evaluation of the expert i to the indicator u_j .

The "typical sort" qualitative and quantitative transformation of the membership function:

$$x(I) = -\lambda p_n(I) \ln p_n(I) \quad (1)$$

Simplify (1):

$$u(I) = -\frac{\ln(m-I)}{\ln(m-1)} \quad (2)$$

I is sorting number, which is given by experts according with the "typical sort". m is transformation parameter, and make $m = j+2$, $I = a_{ij}$ into the formula (3), $h_{ij}(u(a_{ij})) = h_{ij}$ is obtained, $H = (h_{ij})_{k \times n}$ is membership matrix. The "consensus" of index u_j from K experts, is the average degree of knowledge, denoted by $h_j = (h_{1j} + h_{2j} + \dots + h_{kj}) / k$.

Uncertainty generated by the perception known as the "degree of understanding of the blind", namely Q_j :

$$Q_j = \left| \left[\max(h_{1j}, h_{2j}, \dots, h_{kj}) - h_j \right] + \left[\min(h_{1j}, h_{2j}, \dots, h_{kj}) - h_j \right] \right| / 2 \quad (3)$$

To each u_j , overall degree of understanding of k experts about u_j is x_j , $x_j = h_j(1-Q_j)$. Evaluation vector of k experts about u_j is $X = (x_1, x_2, \dots, x_n)$.

Step 3: Normalized

$$w_j = x_j / \sum_{i=1}^m x_i \quad (4)$$

w_j is integrated entropy weight about each index j .

3.2 Fuzzy Mathematic Method to Calculate Results

The key of modeling by using fuzzy mathematical methods is to establish realistic membership functions. Common membership functions are following: a large normal distribution function, partial large Cauchy distribution function, partial small-normal distribution function, partial small Cauchy distribution function and so on. According to the actual situation, selecting the partial large Cauchy distribution function and logarithmic functions as a membership function:

$$f(x) = \begin{cases} [1 + \alpha(x - \beta)^{-2}]^{-1}, & 1 \leq x \leq 3 \\ a \ln x + b, & 3 < x \leq 5 \end{cases} \quad (5)$$

In (6), α, β, a, b are undetermined constants. So, an evaluation object would have an evaluation matrix $x_i = (r)_{1 \times n}$

$$X = \sum_{i=1}^{10} Wx_i \quad (6)$$

According to the formula (6), each supplier's evaluation result is calculated by fuzzy mathematic method, and through the sort of evaluation, the best IT-outsourcing supplier is selected.

4 Analysis of Example

In this section, an example in IT-outsourcing supplier selection is introduced to confirm the algorithm's validity. Suppose a company has four IT-outsourcing suppliers to be selected, we use model in section 3 to select the best IT-outsourcing supplier.

We first selected 30 experts, groups of 10, survey of specific indicators, collected "typical sort" results of "expert opinions" in Table 1.

Table 1. "typical sort" results of "expert opinions"

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
First group	6	9	5	6	3	4	1	2	10	8
second group	8	6	5	9	4	2	2	1	9	7
Third group	5	6	7	7	2	4	3	1	8	7

Using structure entropy method, Calculation results are shown in table 2.

Table 2. Calculation results by using structure entropy

	h_j	Q_j	x_j	W
C1	0.7123	0.0175	0.6998	0.0935
C2	0.6509	0.0482	0.6195	0.0828
C3	0.7647	0.0234	0.7468	0.0998
C4	0.6255	0.0228	0.6112	0.0817
C5	0.9146	0.0009	0.9138	0.1221
C6	0.8982	0.0155	0.8843	0.1182
C7	0.9589	0.0007	0.9582	0.1281
C8	0.9867	0.0066	0.9802	0.1310
C9	0.4418	0.0082	0.4382	0.0586
C10	0.6402	0.0155	0.6302	0.0842

According to indicators, the use of questionnaires, expert scoring, and the evaluation of data are obtained. Evaluation can be divided into five grades {very satisfied, satisfied, more satisfied, not very satisfied, very dissatisfied}, corresponding to 5, 4, 3, 2, 1, data are shown in Table 3.

Table 3. Evaluation data of IT-outsourcing suppliers

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
supplier s1	4	5	2	5	3	4	4	4	5	5
supplier s2	3	3	4	4	2	4	4	3	3	4
supplier s3	3	3	4	3	2	4	4	5	4	4
supplier s4	4	3	2	1	4	2	3	5	5	4

The membership degree of "Very satisfied" is 1, the membership degree of "satisfactory" is 0.8, the membership degree of "very dissatisfied" is 0.01. So, $\alpha = 1.1086$, $\beta = 0.8942$, $a = 0.3915$, $b = 0.3699$. Using formula (5), membership function is:

$$f(x) = \begin{cases} [1 + 1.1086(x - 0.8942)^{-2}]^{-1}, & 1 \leq x \leq 3 \\ 0.3915 \ln x + 0.3699, & 3 < x \leq 5 \end{cases}$$

So, $f(2) = 0.5245$, $f(4) = 0.9126$. Quantify values of table 3 are shown in table 4.

Table 4. Quantify values of evaluation data

	Supplier s1	Supplier s2	Supplier s3	Supplier s4
C1	0.9126	0.8	0.8	0.9126
C2	1	0.8	0.8	0.8
C3	0.5245	0.9126	0.9126	0.5245
C4	1	0.9126	0.8	0.01
C5	0.8	0.5245	0.5245	0.9126
C6	0.9126	0.9126	0.9126	0.5245
C7	0.9126	0.9126	0.9126	0.8
C8	0.9126	0.8	1	1
C9	1	0.8	0.9126	1
C10	1	0.9126	0.9126	0.9126

Using formula (6), evaluation results and sort of each supplier are shown in table 5.

Table 5. Evaluation results and sort of each supplier

Object	supplier s1	supplier s2	supplier s3	supplier s4
Evaluation value	0.8870	0.8240	0.8476	0.7471
Sort	1	3	2	4

So, the best IT-outsourcing supplier is s1.

5 Conclusion

In this paper, the structure entropy method is proposed to solve the problem of IT outsourcing supplier selection. Among them, the index system is a lot of controversial

issues. In practice, for different circumstances, experts need to make certain changes and additions. It is intended to provide a new evaluation idea for enterprise to select IT-outsourcing suppliers. In addition, in order to reduce manual calculation errors and improve the evaluation efficiency, we can program to carry out the calculation process.

Acknowledgments. The author wish to thank the scientific research project of Hubei Provincial Department of education (Grant B20121905), under which the present work was possible.

References

1. Chen, Q.-J., Qi, F.: Research Commentary on Selection of Supplier. *Foreign Economics & Management* (5), 31–37 (2009)
2. Dickson, G.W.: An analysis of vendor selection systems and decisions. *Journal of Purchasing* (2), 5–17 (1996)
3. Huang, J., Deng, Z.-Y.: Chooses IT Foreign Contract Project Supplier's Decision-making Based on AHP. *Value Engineering* 25(9), 73–75 (2006)
4. Zhou, B.-X., Wang, Y.-Q., Zhu, Y.-J.: Application of fuzzy comprehensive evaluation method in the IT outsourcing vendor selection. *China Management Informationization* (5), 3–5 (2006)
5. Yang, B.-J., Hou, L.-W.: Selecting IT Outsourcing Suppliers Based on Nested Logit Model. *Industrial Engineering and Management* 14(1), 71–74 (2009)
6. Tam, C.Y., Tummala, V.M.: An application of the AHP in vendor selection of a telecommunication system. *Management Science* 29, 171–182 (2001)
7. Luo, X.-X., Peng, S.-H.: Research on the Vendor Evaluation and Selection Based On AHP and TOPSIS in Green Supply Chain. *Soft Science* (02), 53–56 (2011)
8. Weber, C., Ellram, A.: Supplier selection using multi-objective programming: a decision support system approach. *International Journal of Physical Distribution & Logistics Management* 23(2), 3–14 (1992)
9. Arunkumar, N., Karunamoorthy, L., Anand, S., Ramesh, B.T.: Linear approach for solving a piecewise linear vendor selection problem of quantity discounts using lexicographic method. *International Journal of Advanced Manufacturing Technology* 28(11), 1254–1260 (2006)
10. Cheng, Q.-Y.: Structure entropy weight method to confirm the weight of evaluating index. *Systems Engineering-Theory & Practice* 30(7), 1225–1228 (2010)
11. Han, Z.-G.: Mathematical modeling method and its application. Higher Education Press, Beijing (2009)

Chinese Web Text Classification Model Based on Manifold Learning

Shengli Shi¹, Zhibin Fu¹, and Jinzhao Li²

¹ College of Information Technology, Hebei Normal University,
Hebei Shijiazhuang 050024, China

² School of Computer Science and Engineering, Hebei University of Technology,
Tianjin 300401, China
shengli10@126.com

Abstract. To study a Chinese web text classification model based on manifold learning. Manifold learning methods can effectively map the high-dimensional web text data into a low dimension space. Reducing the dimension of Chinese web text data can improve the efficiency of the classifying algorithms. In this model, the Chinese web text data are firstly reduced the dimensions with ISOMAP. Then the low-dimensional data are classified with Bayes classifier. The result shows that the executing efficiency is greatly improved and the qualities of classification are guaranteed.

Keywords: Manifold learning, Web text classification, ISOMAP, Dimensionality reduction.

1 Introduction

Web text classification is mainly used in the email message classification, information retrieval services, electronic meeting and other fields, it is an important method of analyzing, organizing and managing Web text data. At present, Chinese internet users are more than any other countries, and there are tremendous Chinese Web sites in internet. So how to better meet the needs of Chinese internet users, providing them with better services is an urgent problem to solve, the study of Chinese Web text classification is necessary. Now the researching of Chinese Web text classification is very hot in Web text classification.

Web text classification is based on text classification and text classification has been widely used. But Web texts are massive, heterogeneous, and semi-structure and dynamic, so the retrieval and classification of web text is much more difficult than normal texts. We need to give the web texts real-time analyze from multiple perspectives, multiple levels. At present, there are several text classification algorithms used broadly. They are k-nearest neighbor algorithm [1] (k-Nearest Neighbor, KNN), naive Bayesian [2] (Naive Bayes, NB), decision trees [3](Decision Tree , DE), Support vector machines [4] (Support Vector Machine, SVM) etc. These classification algorithms can be applied in Web Text classification and there are some automatic text classification system based on these algorithms [5]. But

because Web texts are high-dimensional nonlinear, result in “disasters of dimension”, greatly influence the effect of these classification algorithms. Therefore it is very necessary to reduce the dimension of web text.

There are two type methods to reduce the dimension: linear and nonlinear. Principal component analysis (Principal Component Analysis, PCA) and multi-dimensional scaling transformation (Multi-Dimension Scaling,MDS) are both typical linear dimension reduction methods, they can effectively project linear high dimensional data into low-dimensional space, and has been widely applied. However, these linear methods could not found low-dimensional manifolds hidden in the nonlinear high dimensional data. Manifold learning is a kind of non-linear dimension reduction methods developed in recent years, this kind of methods can effectively map high-dimensional nonlinear data to low-dimensional space. Typical algorithms of Manifold learning include raster mapping method (Isometric Mapping,ISOMAP)[6], and by locally linear embedding (Locally Linear Embedding,LLE)[7], Laplace characteristic operator mapping process (Laplacian eigenmaps) [8] Correction method, cutting space (Local Tangent Space Alignment, LTSA) [9] and so on, all these methods can find the real manifold structure hidden in high-dimensional nonlinear space.

This article presents a Chinese Web Text data classification algorithm based on manifold learning. This algorithm firstly apply ISOMAP Algorithm to Web text data, then select a suitable low-dimensional data based on residuals, and then use the naive Bayesian methods to classify low-dimensional data. The experiment results verify the validity of this algorithm.

2 Introduction of ISOMAP Algorithms and Naive Bayesian Classification Method

2.1 Introduction of ISOMAP the Algorithm

ISOMAP algorithm is a extension of MDS algorithms in high-dimensional nonlinear space. it can effectively map high-dimensional nonlinear data to low-dimensional space. Its main principle is using geodesic distance instead the Euclidean distance in MDS algorithm. ISOMAP algorithm can effectively retain the characteristics of high-dimensional nonlinear data, so the algorithm has won many research on it. So far, more than 20 species improved algorithms based on ISOMAP have been used to many areas of data mining. Set high-dimensional data sets to $x=\{x_1, x_2, x_3, \dots, x_n\in R^D\}$, general steps of ISOMAP algorithm is as follows [6]:

(1) Construct neighborhood. Selecting k nearest neighbors $x_j(i=1,2,3,\dots,N; j=1,2,3,\dots, k)$ for each data point in high-dimensional data set X , getting an adjacency graph G . Usually there are two ways to identify the neighborhood of each point: ϵ -neighborhood and k - nearest neighbor, and taking continental distance between data points as weights of the edges in adjacency graph G .

(2) Estimate the GEODESIC distance. Estimating the GEODESIC distance between any two points in data set x . For nearest neighbor point, directly enter the distances to approximate GEODESIC distance; For non-neighbors, using the shortest path $d_G(i,j)$ between two points in neighborhood graph G to approximate, The shortest

path between two points can be solved using Floyd algorithm or Dijkstra algorithm. Taking the distance matrix as $D_G=\{d_G(i,j)\}$.

(3) Execute MDS Algorithm on Distance matrix D_G , construct d Dimension European space embedding Y , The space maximally maintained the inner geometry of manifolds.

2.2 Naive Bayes (NB) Classification Algorithms

Naive Bayes classification is a simple but effective method of statistical classifications, which are based on the assumption that: in a case with multiple properties, text properties are not related. In other words, the effect of one property of texts for classification has nothing to do with the other properties. Set a_i represent any text, belong to certain b_i which is one text class in $B=\{b_1, b_2, \dots, b_k\}$. The algorithm can be described as follow [2]:

Based on Naive Bayes method expression (1),(2) are true:

$$P(a_i) = \sum_{j=1}^k P(b_j)P(a_i / b_j) \quad (1)$$

$$P(b_j / a_i) = \frac{P(b_j)P(a_i / b_j)}{P(a_i)} \quad (2)$$

(2) When we classify text a_i , Calculates probabilities of all the text in a given a_i according expression (1), a_i will be in the class which has the biggest probability values. That means:

$$a_i \in b_j / P(b_j / a_i) = \max_{j=1}^i \{p(b_j / a_i)\} \quad (3)$$

From (1),(2) and (3) we can say, if text and background of tests are given, the key step of using Naive Bayes method is calculating $P(b_i)$ and $P(a_i/b_j)$. The process of calculating $P(b_i)$ and $P(a_i/b_j)$ is the process of establishing a classification model.

Because Naive Bayes method is based on the text properties' independence, that is assuming the probability of a word occurrence in the text has nothing to do with any other factors in the text, this greatly simplifies the calculation, improve the efficiency. But for the vast majority of real text, this assumption is not valid, so the accuracy will be lower than desired.

3 Chinese Web Text Classification Model Based on Manifold Learning

Web Text classification is in fact a fundamental process of pattern recognition. The process is divided into two parts: training and testing, these two parts are two relatively independent process, the first step establishing classifier by training process, the second step test the effect of the classification system by test set. The input of system is a well classified text corpus, the output of the system is the structure of the classifier, that means the characteristics model of each class. The

testing process is text classifier classify the test texts based on the characteristics model. Test texts are the input to the classifier, the categories of tests are the output.

Due to the High-dimensional characteristics of the Chinese Web text data, the results of the traditional classification method is not ideal, this article presents a Chinese Web Text classification model based on manifold learning. First reduced the dimensions of high dimensional Web text data with manifold learning methods, and then using normal classification algorithm to classify the texts which has reduced dimensions. Chinese Web text classification model based on manifold learning is shown as Fig. 1:

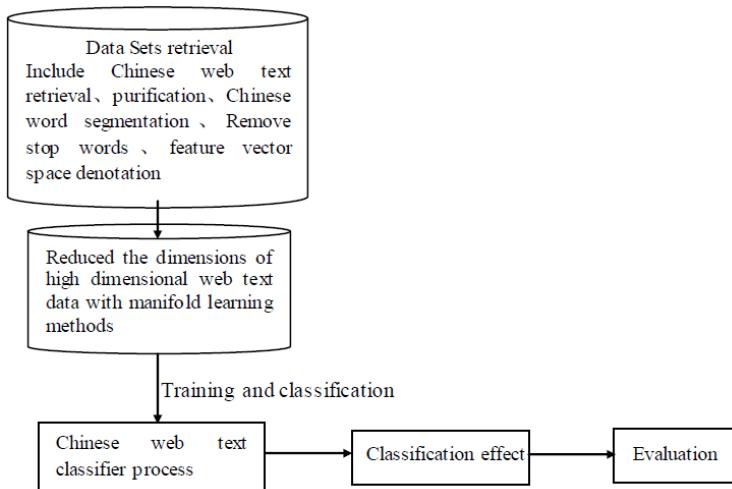


Fig. 1. Chinese web Text classification model based on manifold learning

4 Results of Experiment

In order to verify the validity of the algorithm, This article use a corpus build by Tan Songbo,Wang Yuefen [10,11]. The corpus is divided into two levels, collecting 14150 web text articles. The first level includes 12 categories; the second level includes 60 categories. The Table 1 shows how many articles there are in each category of level one:

Table 1. The count of articles in each category of level one of the corpus

Category	Financial	Sports	Education	Health	Science and technology	Art
Counts	819	2805	808	1406	1040	546
Entertainment	HR	Geographical	Car	Computer	Estate	
1500	608	150	590	2943	935	

For ease of calculation, this article randomly selects 200 texts from each category of the corpus but 150 texts from Geographic category, forming the experimental data which has 2350 texts. There are 47926 characteristic words in the experimental data,

which means the dimension of the data is 47926, so it can be represented as a 2350*47926 matrix. This data set is divided into a training set and a test set, the geographic text set is divided according to the ratio 2:1 of training text and test text. The other category texts are divided according to the ratio 3:1 of training text and test text. Each text category has its category tag. In order to compare the effect of the algorithm before and after dimension reduction, firstly perform the algorithm on the original data set, and compute precision, searching rate, and F1 values and calculates the time consuming; secondly, reduce the dimension of Chinese web text data with ISOMAP algorithm, and respectively execute the classification algorithm when the dimension is 6 or 50, then compute precision, searching rate, F1 value and time consuming. When perform ISOMAP algorithm the neighborhood parameter k should be 5, the classifier is Naive Bayes Classifier. Precision was shown in Fig. 2, the searching rate was shown in Fig. 3, the F1 value was shown in Fig. 4, time consuming on different dimensions was shown in Table 2.

The experimental results tell us that the precision, the searching rate and the F1 values are similar in the raw data and low-dimensional data situation, but the execution efficiency of low-dimensional data is much higher than the original high-dimensional data. So the Chinese web text classification algorithm based on manifold learning can greatly improve classification efficiency and guarantee the classification effect.

Table 2. Comparison of the classification efficiency before and after dimension reduction

The data set tested	original data	k=5,d=6	k=5,d=50
time consuming(seconds)	16154	923	1276

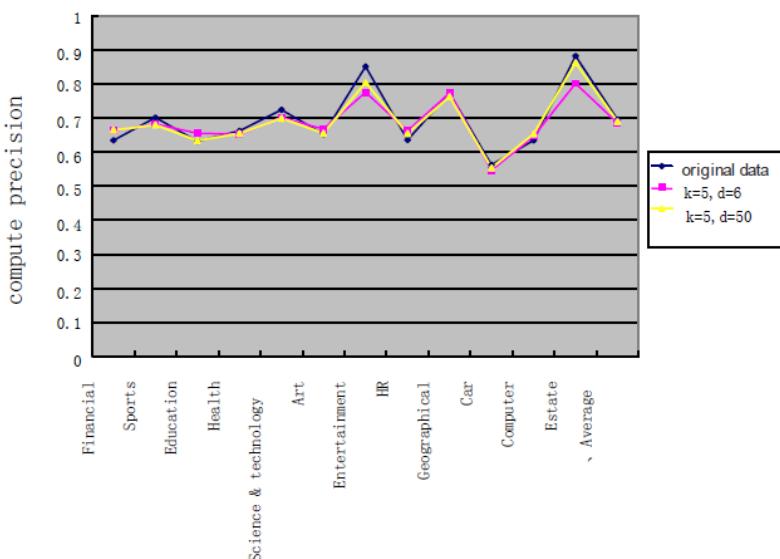


Fig. 2. Comparison of precision before and after dimension reduction

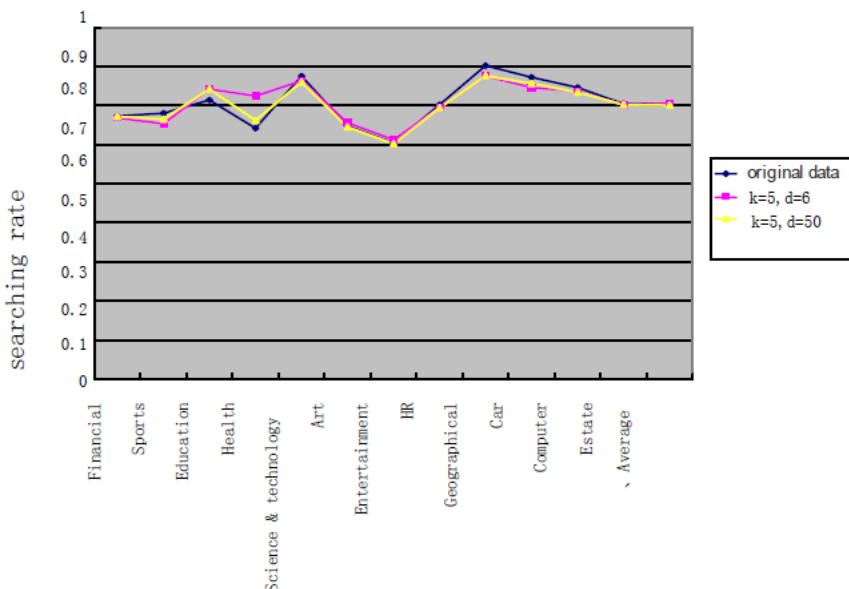


Fig. 3. Comparison of Searching rate before and after dimension reduction

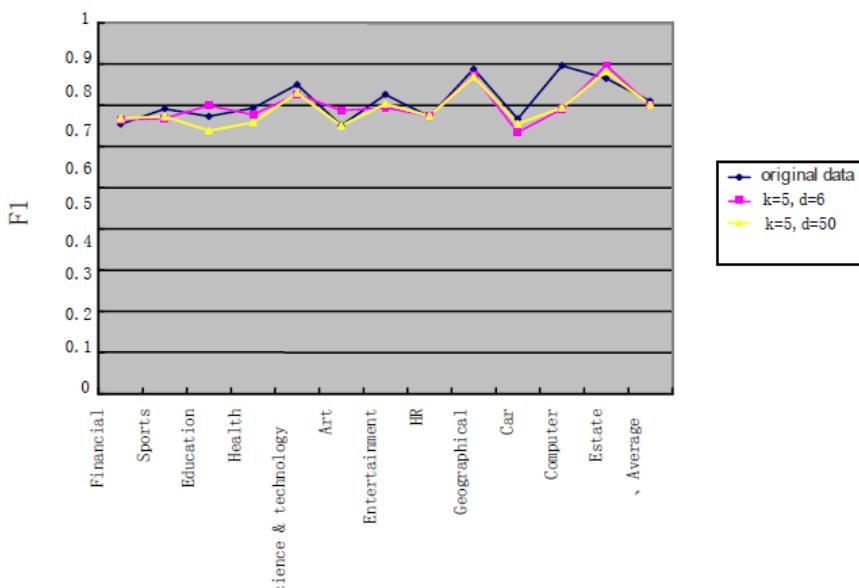


Fig. 4. Comparison of the F1 before and after dimension reduction

5 Conclusion

As the Internet is more and more popular, web texts are growing at exponential rate, which will undoubtedly give the text categorization researchers a serious problem: disaster of dimension. The heterogeneity and complexity of high-dimensional data reduce the classification quality of many classifiers; The traditional feature extraction method for text categorization has been an important dimension reduction methods in the field. While the tedious thresholds determination or mathematical calculations in feature extraction method increase the overall amount of time for text categorization, and dimension reduction algorithm based on manifold learning is a method which has faster speed and better effect. The experiment results show that the algorithm presented in this article can greatly improves the performance of text classification and the premise of the classification is also good.

Acknowledgment. Supported by Natural Science Foundation of Hebei Province of China (project:F2011205032).

References

1. Yu, R.-P.: Research and implementation of Chinese text categorization algorithm. Master degree thesis, Northwestern University Xian (2007)
2. Han, J., Kamber, M.: Data mining concepts and technologies.M. Translations such as Meng Xiaofeng. China machine press, Beijing (2005)
3. Wang, Y.: And based on decision tree K Nearest neighbor algorithms for text categorization research. PhD thesis. Tianjin University, Tianjin (2006)
4. Ye, Z.S.: application in text classification: master's thesis. Harbin Engineering University, Harbin (2006)
5. Hsu, C.W., Lin, C.-J.: A simple decomposition method for support vector machines. Machine Learning 46(23), 291–314 (2002)
6. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
7. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000)
8. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15(6), 1373–1396 (2003)
9. Zhang, Z.Y., Zha, H.Y.: Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM Journal of Scientific Computing 26(1), 313–338 (2005)
10. Tan, S., Wang, Y.P.: Corpus of Chinese text categorization -TanCorpV1.0 (Web version), <http://www.searchforum.org.cn/tansongbo/corpus.htm>
11. Tan, S., et al.: A Novel Refinement Approach for Text Categorization. In: ACM CIKM 2005 (2005)

Parallel Fault Information Mining Using Integrating Neural Classifier

Hongsheng Su

Dept. of Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730050, China
shsen@163.com

Abstract. To overcome the defects of single neural network during fault diagnosis such as low diagnosis precision, long training time and bad generalizing ability, an integrating neural classifier was proposed for fault information mining in the paper. The method fully utilized the advantages of ensemble neural network, i.e., fault information was mined using several neural networks, simultaneously, and the achieved results were weighted average in accordance with their respective properties. To quicken the convergence speeds of neural networks, Kohonen network was used to discretize the decision table and rough set was then used to implement the redundant attributes reduction. In the end, both simulation and experiment indicate that the proposed method is very effective, and has better generalizing ability.

Keywords: Fault information, mining, ensemble neural network.

1 Introduction

In recent years, with the rapid advancement of artificial neural networks (ANN), ANN has been in an increasingly broad application in diverse fields including artificial intelligence (AI), pattern recognition (PR), expert systems (ES), image processing (IP) and process control (PC) as well as many other fields. Due to its prominent characteristics such as self-learning and non-linearity as well as parallel distributed information storage ability, ANN has already been considered to be very suitable to facility fault diagnosis[1], and has been successful application in power systems [2,3], mechanical engineering [4,5], chemical engineering [6,7] and control systems [8,9], and etc. Clearly, to achieve parallel calculation and nonlinear mapping from input space to output space are the most excellent properties of ANN. However, single ANN exposes some problems during applications, say, incredible results, inscrutable behavior, and weak generalizing ability, which restricts its applications farther. Hence, to overcome the flaws and achieve the best results, to train more neural networks and integrate their results is considered a realistic selection [10]. This dramatically improves the generalizing ability of single neural networks. In addition, the handicaps to choose network model is also better removed. Based on it, an ensemble neural classifier is proposed for stream turbine vibration faults diagnosis in the paper. The proposed approach gives diverse weights to the posterior probabilistic

estimations of each network according to their respective properties, and the identification rate is improved by weighted average. Meanwhile, Kohonen network is used to discretize the decision table and rough set is used to implement the redundant attributes reduction, the training speed of the networks is therefore improved, dramatically.

The following sections include Kohonen network based data discretization method, neural networks ensemble principle, practical example and application, conclusions, and etc.

2 Kohonen Network Quantification Method

The diagnostic decision table possibly comprises some redundant characteristics. Therefore, how to identify these redundant characteristics and ignore them while not influence the content of fault diagnosis will be a very significant job. As rough set is considered as a powerful tool to data mining and knowledge discovery, it can be used to simplify the decision table. Due to rough set of its discrete disposal, the continuous attributes values in decision table must be discretely disposed. Considering the excellent characteristics of Kohonen network, such as simple structure, short learning time, high classification precision and strong anti-interference ability, Kohonen network is therefore broadly applied to discretize the decision table, and the algorithm is described as follows[11].

Step 1: Given the largest classification number of Kohonen network I , the set V_i of each attribute value is classified starting from the two classes, classification boundary point $P(j)$ is then achieved.

1. A smaller stochastic variable is assigned to initial weigh vector $W(O)$ of Kohonen network.
2. The distance d_j between input $X_i(t)$ and each output node j can be worked out.
3. The connection weighs of the minimal distance node is adjusted.

Step 2: When classification boundary point is $P(j)$.

1. According to $P(j)$ region partition is implemented, and each region is served as a class, and the variance D_1 inside class and the variance D_2 between the classes are respectively worked out.
2. Set K ($I \geq K \geq 2$) is the pre-classification number, while $K \geq I$, turn to step 1, while $K < I$, the classification is terminated. Hence, $K-1$ is the largest classification number of the attribute set.
3. For every attribute set V_i , K is selected as the best classification number only because it can let D_1 up to the minimum and D_2 up to the maximum.

Step 3:

1. According to the best classification of each V_i , the best classification boundary points may be gained.
2. Region partition is performed and marked.
3. The discretized decision table is acquired, finally.

The largest classification number I in the algorithm above can be confirmed in advance based on the samples sorts, or automatically set by program. Though, the selection of I doesn't influence the disposal of the results.

3 Principle of Ensemble Neural Network

To overcome the flaws of single neural networks, an optional method is to train more neural networks and integrate their results. Figure 1 is the ensemble principle.

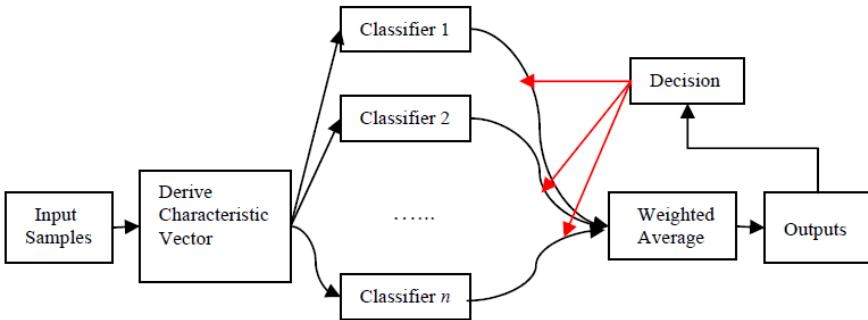


Fig. 1. The principle diagram of neural networks integration

Let classification objects have c patterns, that is $\omega=(\omega_1, \omega_2, \dots, \omega_c)$. Assume that M neural classifiers are used to integrate in Fig.1. Let input samples be X , the true posterior probability distribution be $q(X)=[p(\omega_1|X), p(\omega_2|X), \dots, p(\omega_c|X)]$. The estimation of the posterior probability of the i^{th} neural classifier may be described by

$$q_i(X) = [\hat{p}_i(\omega_1|X), \hat{p}_i(\omega_2|X), \dots, \hat{p}_i(\omega_c|X)], i=1,2,\dots,M. \quad (1)$$

Due to insufficient samples and incomplete training, the error $\varepsilon_{ij}(X)$ between the true posterior probability $p_i(\omega_j|X)$ and the estimated posterior probability $\hat{p}_i(\omega_j|X)$ is existed and expressed as $\varepsilon_{ij}(X)=p_i(\omega_j|X)-\hat{p}_i(\omega_j|X)$, the error rate is higher than one of Bayesian classer. Hence, the correctness of the posterior probability should be improved. To resolve the problem, we apply weighted average sum of each posterior probability to reduce the estimated error. Considering the differences among diverse networks and their relativities, diverse weights are endowed to diverse networks. Let the weight value of each classifier be $w_i, i=1,2,\dots,M$, then

$$\hat{q}(X) = \sum_{i=1}^M w_i q_i(X) \quad (2)$$

The principle to select the weights to let error $e = \frac{1}{2} \sum \| q(X) - \hat{q}(X) \|_2^2$ lower than the minimum ε , this means

$$\frac{dE\{\|q(\mathbf{X}) - \hat{q}(\mathbf{X})\|^2\}}{dw_i} = 0, i = 1, 2, \dots, M. \quad (3)$$

$$E\{\|q(X) - \hat{q}(X)\|^2\} = \sum_{j=1}^c E\{(p(\omega_j | X) - \sum_{i=1}^M w_i \hat{p}_i(\omega_j | X))^2\}. \quad (4)$$

Therefore,

$$\begin{aligned} & \sum_{i=1}^M w_i \sum_{j=1}^c E\{(\hat{p}_i(\omega_j | X) \hat{p}_k(\omega_j | X))\} \\ &= \sum_{j=1}^c E\{p(\omega_j | X) \hat{p}_k(\omega_j | X)\} \\ & k = 1, 2, \dots, M \end{aligned} \quad (5)$$

Let

$$\begin{aligned} i_{ik} &= \sum_{j=1}^c E\{\hat{p}_i(\omega_j | X) \hat{p}_k(\omega_j | X)\}; \\ o_k &= \sum_{j=1}^c E\{p(\omega_j | X) \hat{p}_k(\omega_j | X)\}. \\ \mathbf{W} &= \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_M \end{bmatrix}, \quad \mathbf{I} = \begin{bmatrix} i_{11} & i_{12} & \dots & i_{1M} \\ i_{21} & i_{22} & \dots & i_{2M} \\ \dots & \dots & \dots & \dots \\ i_{M1} & i_{M2} & \dots & i_{MM} \end{bmatrix}, \quad \mathbf{O} = \begin{bmatrix} o_1 \\ o_2 \\ \dots \\ o_M \end{bmatrix}. \end{aligned} \quad (6)$$

Then, the formula (5) may be written as

$$IW = O \quad (7)$$

$$W = I^{-1}O \quad (8)$$

And so,

$$E\{\|q(X) - \hat{q}(X)\|^2\} = \sum_{j=1}^c E\{(p^2(\omega_j | X)) - O^T I^{-1} O\}. \quad (9)$$

Let N training samples be x_1, x_2, \dots, x_N , then

$$\hat{i}_{ik} = \frac{1}{CN} \sum_{n=1}^N \sum_{j=1}^c \hat{p}_i(\omega_j | x_n) \hat{p}_k(\omega_j | x_n) \quad (10)$$

$$\hat{o}_k = \frac{1}{CN} \sum_{n=1}^N \sum_{j=1}^c p(\omega_j | x_n) \hat{p}_k(\omega_j | x_n) \quad (11)$$

According to the equations (8), (10) and (11), we may work out W . While the complexity of the resolved problem is lower, the calculation of the reverse matrix is no difficult, however, and the complexity of the resolved problem is larger, the calculation of the reverse matrix is influenced by data noise coming from individual network easily, which results in a homologous fall in the generalized ability of integration neural networks. At the moment, we may pick up some good networks with high precision to integrate and delete some bad networks with low precision so as to reduce the complexity of the problem.

4 Application

Table 1 is some familiar fault sources and fault symptoms information of stream turbine vibration faults [12], where ten kinds of typical faults in rotation machines are selected as outputs of neural networks, and the score peak energy values in the range of nine frequency ranges are used to act as input characteristic vector, in this way, a sheet of two-dimensional decision table is formed.

Table 1. Fault sources and symptoms table of stream turbine

Sample(D)	a	b	c	d	e	f	g	h	i
1	0.00	0.00	0.00	0.00	0.90	0.05	0.05	0.00	0.00
2	0.00	0.30	0.10	0.60	0.00	0.00	0.00	0.00	0.10
3	0.00	0.00	0.00	0.00	0.40	0.50	0.10	0.00	0.00
4	0.10	0.80	0.00	0.10	0.00	0.00	0.00	0.00	0.00
5	0.10	0.10	0.10	0.10	0.20	0.10	0.10	0.10	0.10
6	0.00	0.00	0.00	0.00	0.20	0.15	0.40	0.00	0.25
7	0.00	0.00	0.10	0.90	0.00	0.00	0.00	0.00	0.00
8	0.00	0.30	0.10	0.60	0.00	0.00	0.00	0.00	0.00
9	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
10	0.00	0.00	0.00	0.00	0.00	0.80	0.20	0.00	0.00

Let input characteristic vector be $X=\{a, b, c, d, e, f, g, h, i\}$, and output vector be $Y=\{D\}$. According to quantification method of Kohonen network, a quantitative decision table is gained as shown in Table 2.

Table 2. Quantitative decision table

Sample(D)	a	b	c	d	e	f	g	h	i
1	0	0	0	0	1	0	0	0	0
2	0	1	1	1	0	0	0	0	1
3	0	0	0	0	1	1	0	0	0
4	1	1	0	1	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1
6	0	0	0	0	1	1	1	0	1
7	0	0	1	1	0	0	0	0	0
8	0	1	1	1	0	0	0	0	0
9	1	0	0	0	0	0	0	1	0
10	0	0	0	0	0	1	1	0	0

After rough set reduction[13], $\{b,e,i\}$ is core attribute set, we therefore can get reduction attribute sets $\{a,b,e,f,i\}, \{a,b,e,g,i\}, \{b,c,e,f,i\}$, and $\{b,c,e,g,i\}$. Here the attribute set $X'=\{a,b,e,f,i\}$ is selected as conditional attributes and D serves for decision attribute, we can get a diagnosis decision table easily. Now let $X'=\{a,b,e,f,i\}$ serve for the inputs of BP network, and the decision attribute D be the outputs of the BP network, a simplified decision table is gained as shown Table 3.

Table 3. Simplified decision table

D	a	b	e	f	i
1	0	0	1	0	0
2	0	1	0	0	1
3	0	0	1	1	0
4	1	1	0	0	0
5	1	1	1	1	1
6	0	0	1	1	1
7	0	0	0	0	0
8	0	1	0	0	0
9	1	0	0	0	0
10	0	0	0	1	0

If the condition attributes in Table 3 are used to serve for the inputs of BP network, and the decision attribute is the outputs of the network, and the samples in Table 3 are also used to train the network. Then the trained-well network may be applied to implement fault diagnosis. Presently, to select the suitable structure of neural network still hasn't theoretical instructions, according to prior experience, the sum of the nodes in hidden layer shouldn't be smaller than the sum of the nodes in input layer. For convenient analysis, we select three neural networks to integrate for faults diagnosis in stream turbine, and the structures of these neural networks are selected as 5-11-10. The transfer function of the neurons in middle layer selects S type tangent

function *tansig*, and one in output layer is S type logarithm function *logsig*. The reason to do this is that outputs of the functions is within the scope of 0 to 1, output requirements of the networks can just be well satisfied. Learning function is *trainlm* function. Let learning times of the networks be 1000, training aim be 0.001, and learning rate equal one, we can get the learning process of the three networks below.

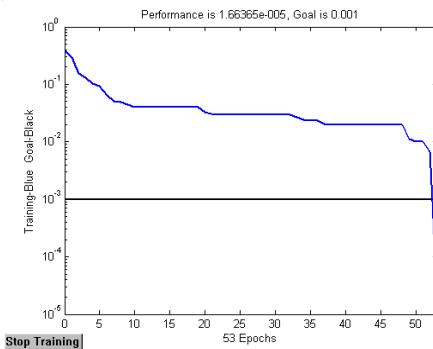


Fig. 2. Training process of the network 1

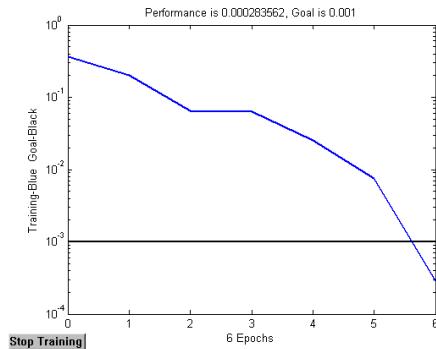


Fig. 3. Training process of the network 2

The samples in Table 3 are respectively inputted to the above networks 1 and 2 and 3, according to their outputs and (8), (10), (11), we can get $W = [0.9024, -0.0581, 0.0456]$.

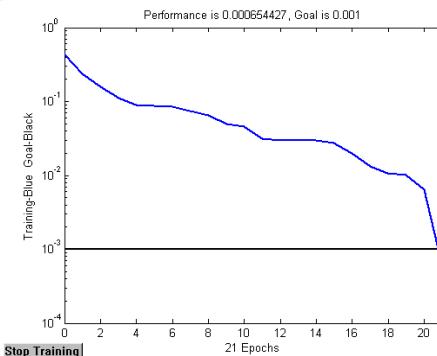


Fig. 4. Training process of the network 3

Since $w_1+w_2+w_3=1.0061\approx 1$, the constraint condition of the equation holds. The following is diagnostic example.

Example: Fault pattern is shown in Table 4, they are respectively inputted to the neural network 1, 2 and 3, and the output results of the networks are shown by Y_1 and Y_2 and Y_3 , respectively.

Table 4. Fault pattern table

a	b	c	d	e	f	g	h	i
0.39	0.07	0.00	0.06	0.00	0.13	0.00	0.00	0.35

$Y_1=(0, 0, 0, 0, 0, 0, 0, 0, 1, 0);$
 $Y_2=(0.0003, 0.0013, 0.0012, 0.0019, 0.0001, 0.0380, 0.0005, 0.0002, 0.7775, 0.0036);$
 $Y_3=(0.0001, 0, 0, 0.0052, 0, 0, 0, 0, 0.9164, 0).$

Clearly, the diagnostic result of the neural network 1 is fault type 9 according to Y_1 , and the diagnostic precession is 100%. From Y_2 , we can also get the diagnostic result 9, but the diagnostic precession is only 77.75%. And from Y_3 , we get the same result in 91.64 %.

Now we integrate their diagnostic results expressed by Y .

$$Y=w_1 \times Y_1 + w_2 \times Y_2 + w_3 \times Y_3 = (0, 0, 0, 0.0003, 0.0002, 0.0025, 0, 0, 0.9860, 0.0002)$$

According to Y , we can know that the diagnostic result is also fault type 9 easily, and the reliability is up to 98.6. Hence, ensemble neural network improves the reliability of the diagnostic result. In addition, since w_1 is very greater than w_2 and w_3 , then the second and third networks may be eliminated.

This is totally similiar to the result in [12], But here the applied method is fully differencent. In the text above we only analyse the ensemble of three neural classifiers, and their types are also homogeneous, actually, we also apply more neural classifiers to integrate and their types may be heterogeneous.

5 Conclusion

The proposed method in this paper can effectively integrate the diagnosis results of more neural networks, and the correctness of fault diagnosis therefore is improved greatly. In addition, while the structure of the networks is complex, those network classifiers with low diagnosis precision may be eliminated, thus the complexity of the algorithm may be reduced. Both simulation and experiment indicate the proposed method is extremely effective.

References

1. Yu, H.J.: Intelligent Diagnosis Based on Neural Networks, 1st edn. Metallurgy Industry Publication House, Beijing (2000)
2. Su, H.S., Li, Q.Z.: Application Study of Composite Neural Networks in Fault Diagnosis Systems Based on GA Learning. Computer Science 32, 255–256 (2005)
3. Su, H., Dang, J.: Composite Neural Nnetworks Applied for Identifying Stream Turbine Vibration Fault Based on Improved PSO. WSEAS Trans. on Circuits and Systems 5, 617–624 (2006)
4. Su, H.: Main Converter Fault Diagnosis for Power Locomotive Based on PSO-BP Neural Networks. Advanced Materials Research 267, 271–276 (2011)

5. Wang, Z., Wang, T., Xu, Y.: Multipile Symptom Fuzzy Fault Diagnosis for Stream Turbine Vibration Based on CLIPS. *China Electric Power* 41, 67–70 (2008)
6. Hoskins, J., Kaliyur, K., Himmelblau, D.: Fault Diagnosis in Complex Chemical Plants Using Artificial Neural Networks. *AIChE Journal* 11, 137–141 (1991)
7. Su, H., Wang, X., Chen, H.: An ANFIS Based Fuzzy Synthesis Judgment for Transformer Fault Diagnosis. In: Yu, W., He, H., Zhang, N. (eds.) *ISNN 2009, Part II. LNCS*, vol. 5552, pp. 172–181. Springer, Heidelberg (2009)
8. Narendra, K.S., Mukhopadhyay, S.: Adaptive Control of Nonlinear Multivariable Systems Using Neural Networks. *Neural Networks* 7, 737–742 (1994)
9. Su, H.: Turbo-Generator Vibration Fault Diagnosis Based on QPSO-BP Neural Networks. In: Proceedings of Power and Energy Engineering Conference (PEEC 2010), pp. 600–603. Sci Res Publish, Irvine (2010)
10. Lam, L.: Optimal Combinations of Pattern Classifiers. *Pattern Recognition Letters* 16, 945–954 (1995)
11. Pan, L., Zhang, L., Zhang, J.: Application of Rough Sets in Image Feature Selection. *Journal of Data Acquisition & Processing* 17, 42–49 (2002)
12. Zhang, Y., Su, H.: Turbo-Generator Vibration Fault Diagnosis Based on PSO-BP Neural Networks. *WSEAS Trans. on Sytems and Control* 5, 37–47 (2010)
13. Zeng, H.: *Intelligence Computation*. Chongqing University Press, Chongqing (2004)

Research of Knowledge Integration Based on Semantic Web for Drilling Risk Management

Yingzhuo Xu

Institute of Computer, Xi'an Shiyou University, Xi'an, China
yzhxu@xsysu.edu.cn

Abstract. Focusing on the practical difficulties of sharing and application risk management knowledge resources due to the distributed and heterogeneous characteristics in today's drilling industry, integration methods as well as key technology and retrieval applications of drilling risk management knowledge were studied by introducing the semantic web technology. Based on the field characteristics and requirements for the sharing of drilling risk knowledge, ontology was selected as the knowledge representation method, establishing a drilling risk knowledge ontology. A framework system supporting the risk knowledge integration based on semantic web was constructed, thereby realizing seamless knowledge integration on the semantics level. On the basis of the ontology model, a prototype of knowledge semantic retrieval system was developed. Finally, an example of knowledge retrieval analysis was provided, with results demonstrating that the retrieval method established in the new system can improve the recall ratio more effectively than the traditional method of keyword retrieval.

Keywords: drilling risk management, knowledge integration, semantic web, ontology, knowledge retrieval.

1 Introduction

Drilling risk management is a very complicated system project which includes the process of risk identification, assessment, prevention and treatment. The process, based on previous experience and knowledge in risk management, must be accomplished by collaboration of multi-domain experts and technicians from such fields as geology and drilling, etc. The drilling industry has accumulated a vast amount of knowledge in long-term risk management, however, because these knowledge resources are often spread out amongst different departments and different business systems, because there are great differences in knowledge representation and storage method, and the semantic description of the same knowledge is not uniform, it is difficult to understand, share and reuse this knowledge, which leads to difficulties for risk management personnel in obtaining knowledge and the squandering of knowledge resources. In order to address these issues, it is necessary to effectively integrate risk management knowledge in the drilling industry and to provide a strong solution to the heterogeneity of knowledge sources problem.

To this end, based on semantic Web technology, this paper puts forward a set of system framework supporting risk management knowledge integration, organically integrating the distributive and heterogeneous knowledge of the drilling industry, eliminating the “semantic gap” in knowledge, realizing knowledge semantic retrieval, and providing a support platform for realizing the knowledge sharing and reuse.

2 Framework Design of Knowledge Integration

In order to realize the demand of knowledge sharing in drilling industry, this paper has designed a framework of knowledge integration based on semantic Web for drilling risk management, as shown in Fig. 1.

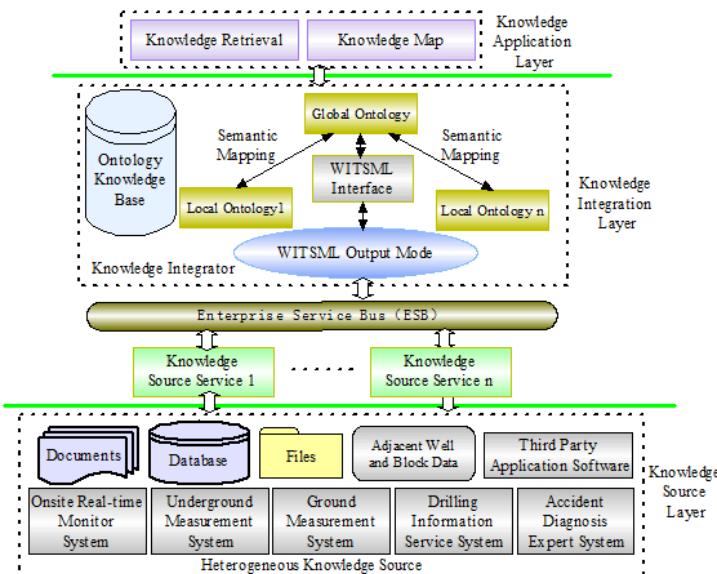


Fig. 1. Framework of knowledge integration based on semantic Web for drilling risk management

Knowledge source layer: it contains all kinds of knowledge resources from existing real-time monitor systems distributed in each department and drilling site, Measurement While Drilling system, ground measurement (comprehensive logging and mud logging) system, drilling engineering service software, accident diagnosis expert system as well as site information of adjacent Wells and Blocks. Their storage forms include all types (Access, SQL Server and so on.) database, documents (Word, Excel, etc.) and graphics/image files.

Knowledge integration layer: it consists of Enterprise Service Bus (ESB) and knowledge integrator, as to the heterogeneous environment of the whole oil company, realize the integration of distributed, heterogeneous risk knowledge by semantic Web

technology, forming uniformed structured knowledge and then storing knowledge base, and ultimately provide them to users though knowledge application layer.

1) ESB: set Enterprise Service Bus between heterogeneous knowledge sources and knowledge integrator, draw the needed knowledge source services from different knowledge sources using ESB technology, describe these services by Web service and then connect to ESB. The services communicate with ESB through standard SOAP message. Each service component of ESB corresponds to a kind of integration solution for knowledge source [1], describing the data transmission mode and format conversion, etc.; at the same time use business events (like data inserting) drive mode to transmit timely new information to knowledge integration system according to redefined ESB configuration, improving the efficiency and automation of integration.

2) Knowledge integrator: realize the structure and semantic description, along with integration of knowledge, thereby to eliminate heterogeneity of the knowledge.

a) Description of risk knowledge structure: Different knowledge sources can have different knowledge patterns (such as ER mode, OO mode, etc.), and the integration of multiple knowledge sources should guarantee the autonomy of each knowledge source, that is the knowledge resource based on the original knowledge model can normally use. Therefore, this paper, based on the international general drilling information standard WITSML (Wellsite Information Transfer Standard Markup Language), uses WITSML mode as output's, and realizes interaction between modes by WITSML interface. WITSML is the well site information transmission markup language based on XML [2]; its purpose is to realize seamless data exchange between oil companies and well site. Coding and describing all kinds of knowledge based on a unified standard and knowledge pattern by WITSML, and it can shield the private mode of knowledge sources, realize seamless knowledge connection.

b) Semantic description and integration of risk knowledge: By using ontology which is a core technology of semantic Web [3], build local ontology based on each knowledge source and merge into unified semantic model “global ontology”. And then standardize description of knowledge mode by the global ontology. Consequently, it realizes the semantics integration of knowledge sources.

Knowledge application layer: provide knowledge application services (such as knowledge retrieval) to users. Realize transparent access and use to drilling risk knowledge through a unified access interface.

3 Key Technology to Realize Knowledge Integration

3.1 Representation of Risk Management Knowledge

Knowledge semantic heterogeneity can lead to ambiguities when users understand knowledge in the same territory [4], so to solve the semantic heterogeneity problem effectively, the knowledge representation method must be reasonable. Therefore, this paper proposes a knowledge representation model based on ontology for drilling risk management- namely, Drilling Risk Knowledge Ontology (hereinafter referred to as DRKO) Model. DRKO is a concept system classifying and describing all the drilling

risk knowledge resources, it assures the recognized concepts in the field, provides common understanding of domain knowledge. DRKO can be described as a formal five factors group{C,A^C,R,A^R,H}[5]. Among them, C represents a concept set, A^C represents attribute sets of each concept, R represents sets of semantics relationships between concepts, A^R represents attribute sets of each relationship, H represents hierarchical relationships between concepts. Take "underground complex situation and accident diagnosis ontology" as an example:

$$\text{Accident_DRKO} = \{C_a, A^C, R_a, A^R, H_a\}$$

Among them, Ca = {drilling accident, complex situation & accident, circulation loss, complex situation, sticking, adhesive sticking, collapse sticking, drill string stilling, drilling tool touch sticking,.....}; Ra = {result (drill string stilling, adhesive sticking), reason (drilling tool touch sticking, collapse sticking),}; Ha = {{drilling accident, complex situation & accident}, {circulation loss, complex situation}, {sticking, drilling accident}, {adhesive sticking, sticking},..... }

3.2 Construction of Knowledge Ontology

According to the formalism description of knowledge ontology above, the ontology's construction step is as follows: (1)determine the domain scope of ontology; (2)determine concepts in the field and create concept classes; (3)determine the concept's attributes and create their classes; (4)establish the semantic relation between concepts;(5)establish the semantic relation between attributes; (6)establish concept instances; (7)establish an ontology using established concepts, attributes, relations and instances.

Take "drilling accident diagnosis" in drilling risk management as an example. Through the thorough investigation and analysis, determine the important concepts and relationships in the field as follows: drilling accident diagnosis includes the judgment and analysis of accident type, causes and symptoms, and drilling accident includes many types such as sticking, well blowout, etc. Each kind of accident diagnosis can be seen as classification of one class in the domain ontology. The accident cause involves geological strata, engineering and drilling fluid, etc.; Geological strata information includes stratum type, rock types, etc.; Drilling engineering includes drilling process and well structure; Drilling fluid information includes drilling fluid types and performance. Drilling accidents is closely related to these factors, and the main symptoms include hanging load, torque, pump pressure, machinery drilling rate, etc. The anomalous change of these parameters indicates that drilling accidents will appear. Limited by the length of this paper, Fig. 2 only gives a part of knowledge ontology model for drilling accident diagnosis, in which the rounded rectangle represents attributes, and rectangle represents concept entity.

The coding realization of the ontology uses OWL ontology description language and Protégé which is an ontology construction tool [6], the method is as follows:

1) Construction of concept class in field: use Jena API [6], by calling ModelFactory. CreateOntologyModel method, build an instance of the ontology model class-OntModel, then by calling createClass method of the instance, build concrete instances of each concept class OntClass to obtain all the concepts in the field.

2) Construction of concept's attributes and attribute classes: by calling `createObjectProperty` method of `OntModel` class instance built in step 1), build instances of each attribute class `ObjectProperty` to obtain attributes of all concepts.

3) Construction of semantic or hierarchical relationships between concepts: build the relationships by using `addSubClass` method or `addSuperClass` method of the concept class-`OntClass`.

4) Construction of semantic or hierarchical relationships between attributes: build by using `addSuperProperty` method of the attribute class-`ObjectProperty`. Further more, domain and range of attribute should also be set, and they are set through `addDomain` and `addRange` method of the attribute class-`ObjectProperty`.

5) Construction of concept instance: build instances of conception ontology class by calling `createIndividual` method of the ontology model class-`OntModel`, and building instances of each instance class-`Individual`.

6) Construction of ontology: build ontology by using the built concepts, attributes, relationships and instances, build instances of the ontology class `Ontology` by calling `createOntology` method of `OntModel` instance, and set each additional attribute by calling `addProperty` method of ontology class-`Ontology`.

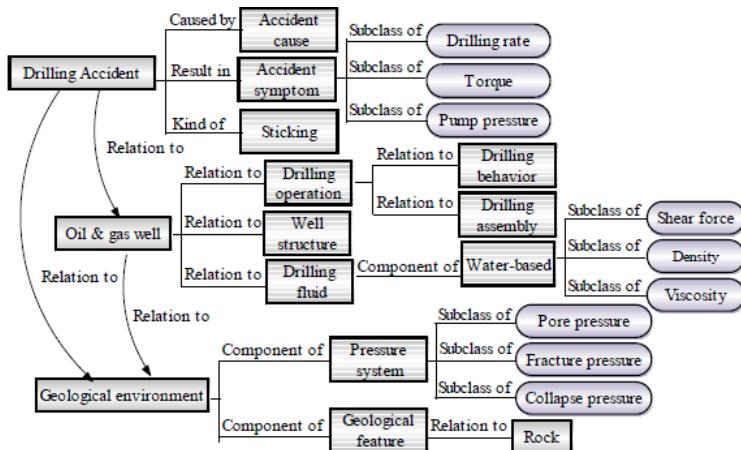


Fig. 2. Knowledge ontology model for drilling accident diagnosis

3.3 Knowledge Integration Based on Ontology

In order to realize the integration of distributed and heterogeneous drilling risk knowledge, this paper has designed a knowledge integrator based on ontology (as shown in Fig. 1). Consistently expressing the semantics of risk management knowledge by knowledge ontology, the integrator translates the knowledge integration problem into the integration of knowledge ontology, through the integrated risk knowledge global ontology, realizing optimization and consolidation of all risk management knowledge. The basic realizing idea is first to build a local ontology to describe semantics for knowledge resources of each business department

(such as geology), and then build a risk domain global ontology in its upper level as well as mapping relationship (namely ontology mapping) between local and global ontology, and integrate knowledge using semantic mapping rules.

1) Construction of local ontology: namely build knowledge ontology for each knowledge source. Through the comprehensive analysis of each knowledge source, extract its knowledge pattern to form WITSML documents, confirm the specific terms and relationship of each knowledge source, and then build local ontology based on each knowledge source using the referred method in section 3.2.

2) Construction of global ontology: Global ontology (namely sharing vocabulary library) is the set of common concepts, characters and objective relationship for describing all the objects in drilling risk management. Since each local ontology of knowledge sources can only form consistent common understanding within the knowledge source, when integrating these knowledge sources, these local ontology can not form consistent common understanding, so we must merge them to build a global ontology, form some mapping relationship between local and global ontology, realizing semantic integration of knowledge sources through the global ontology.

Construction method of the global ontology is first to conduct complete analysis of all knowledge sources, determine the classes, attributes and relationships contained in global ontology according to the knowledge need of risk management, and then abstract the unified class, properties and relationships from each local ontology through ontology mapping, eliminating semantic heterogeneity.

3) Ontology mapping: it defines mapping relationships between local and global ontology, so that the concept in local ontology can map into the same semantic concept in global ontology, eliminating semantic heterogeneity between local ontology. The construction of mapping relationship mainly uses the semantic mechanism of OWL language to realize. The semantic heterogeneity of drilling risk management knowledge is mainly name heterogeneity, including "different name same meaning" and "same name different meaning".(1)Different name same meaning: refers to different knowledge sources use many terms to represent the same concept or attribute, for example, well collapse is also called collapse. This type of heterogeneity is described through equivalent relationship-equivalentClass, equivalentProperty and sameAs, and set by addEquivalentClass of the concept class OntClass. (2)Same name different meaning: refers to different objects use the same name in the same knowledge source. And it is described using OWL semantic tag differentFrom and allDifferent, set by addDisjointWith method of OntClass.

3.4 Knowledge Semantic Retrieval Based on Ontology for Risk Management

Traditional retrieval technology based on keywords uses simple word matching rules [7], therefore, only retrieval words in the retrieval objects may be retrieved, and similar semantic concepts to the query request cannot be retrieved due to using different words, which greatly reduces knowledge sharing. This paper adopts a semantic retrieval model based on ontology, that is, to expand user's retrieval requests from concept level, and retrieve knowledge by the semantic extension (including synonymous, upper and lower expansion) in order to improve the recall ratio.

The realization of retrieval process is: (1) Input query requests from user interface. (2) Extract concepts from the request by the pretreatment module using word splitting and filtering technology, forming a retrieval concept set called as user concept space, and then map them into ontology concept space by XML mapping configuration technology. (3) The semantic reasoning module conducts semantic matching by calculating the semantic similarity between concepts above, finding classes related to concepts, and then conduct reasoning for querying expansion according to class instances obtained by matching the query requests, as well as their corresponding relationship in the ontology knowledge base, getting retrieval keywords after expansion. At last, the module conducts retrieval operation. (4) Return retrieval results to users by the result dispose module after merging and sorting by similarity.

The semantic reasoning module is the core of ontology-based semantic retrieval. It is a reasoning machine consisting of a group of programs with reasoning strategy [8], using the rule-based reasoning engine provided by Jena [9], combining ontology of the ontology base, calculating the semantic similarity among concepts to match, complete retrieval tasks. The similarity calculation is as follows[10]: Suppose that the semantic similarity between user concept A and ontology concept B is $\text{Sim}(A, B)$, if the relation of both A and B is: (1) Synonymous, so $\text{Sim}(A, B)=1$; (2) Inheritance relation, $\text{Sim}(A, B)=1/2$; (3) Part-whole relation, $\text{Sim}(A, B)=2/3$; (4) Indirectly related by n concepts, $\text{Sim}(A, B)=1/(n+1)$; (5) Antisense or disparate, $\text{Sim}(A, B)=0$.

Below emphatically describe the expansion reasoning realization of the reasoning machine in the aspect of "Synonymous" and "Upper and Lower" relationship of the ontology model. "Synonymous" relationship defines the equivalency between A and B[7], when A appears in query conditions, B can be used as its replacement. "Synonymous" is defined by owl:equivalentClass, owl:sameAs. "Upper and Lower" relationship defines upper and lower (namely inclusion) relation among concepts [7], when retrieving a concept, its epigynous and hypogynous concepts also join in the query conditions. "Upper and Lower" relationship is defined by rdfs:SubClassOf. Suppose the query condition, after handled by splitting words, is the set $A=\{a_1, a_2, \dots, a_n\}$, the concept extension set is E, then semantic reasoning algorithm is as follows:

Step 1. Initialize the set $E = A$.

Step 2. As to each element a_i in set A, combining the ontology knowledge base, conduct the following operation: (1) find out all entities that have the relationship of owl: sameAs and owl:equivalent with a_i , join the set E and calculate their similarity. (2) Judge the category of a_i , if a_i is a class or an attribute, and then find out its upper entity and lower entity respectively to join the set E and calculate their similarity.

Step 3. Conduct knowledge retrieval according to the extended key word set E.

4 Realization of the Knowledge Integration Framework

Taking Eclipse and Java JDK as the development platform and using Protégé and ontology interpretation tool Jena, the knowledge integration framework for drilling risk management is realized. Meanwhile, the prototype of knowledge retrieval system

by Jena reasoning machine is developed. Through the semantic expansion reasoning, it realizes the integration retrieval of scattered and heterogeneous knowledge sources.

The following is a concrete example describing the knowledge retrieval process. When taking "adhesive sticking" as a retrieval condition, through semantic expansion, the reasoning machine will implement matching with its synonymous concepts (differential pressure sticking), upper concepts(sticking) and brother concepts(balling up sticking, .etc.), getting the expanded keyword set{adhesive sticking, differential pressure sticking, sticking, balling up sticking, ... }, and then conduct retrieval operation by the search engine. So, the retrieval system returns not only "adhesive sticking" information (including reasons and treatment measures, etc.), but also the accident information of synonymous, upper and brother concept, improving the recall rate of the system and making query results accord with user's retrieval intentions.

5 Conclusion

This paper put forward an ontology-based knowledge representation method for drilling risk management, established risk knowledge ontology, realized integration of risk knowledge by semantic Web, shielding distributivity and heterogeneity of knowledge resources, truly achieving seamless integration of knowledge resources on the semantic level, thereby provided the conditions for effective utilization of knowledge. On this basis, an ontology-based prototype system for knowledge semantic retrieval has been developed. The knowledge recall ratio can be improved using the system, obtaining a greater extent of knowledge sharing and reuse.

Acknowledgment. We are grateful to the National Natural Science Foundation China and the Special Fund Project of Department of Education of Shaanxi Province for their financial supported under Grant No.51074125, No.11JK0790 and No.11JK0793 for this paper.

References

1. Gu, W., Xia, G., You, W.: Method of enterprise knowledge integration based on semantic web. *Journal of Beijing University of Aeronautics and Astronautics* 33, 243–247 (2007)
2. Kirkman, M.A., Symmonds, M.E., Mokhtar, A.: Wellsite Information Transfer Standard Markup Language. In: *WISML, an Update*, SPE 84066, pp. 278–281 (2003)
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American* 284, 34–43 (2001)
4. Oberle, D., Staab, S., Studer, R.: Supporting application development in the Semantic Web. *ACM Transactions on Internet Technology* 5, 328–358 (2005)
5. Brandt, S.C., Morbach, J., Miatidis, M., et al.: An ontology-based approach to knowledge management in design processes. *Computers & Chemical Engineering* 32, 320–324 (2008)
6. Wang, H., Han, H.: Web Ontology Language OWL and Reasoning. *Coal Technology* 29, 217–218 (2010)
7. Liu, Q., Shang, Y.: Research on Semantic Knowledge Retrieval Based on Domain Ontology. *Library and Information Service* 55, 61–65 (2011)

8. Wu, Z., Sun, X., Li, X.: Application of Ontology Based Reasoning in Knowledge Base of Industry Information. *Computer Engineering* 34, 59–61 (2008)
9. Chai, L., He, F.: Research on Jena-Based and Ontology Reasoning. *Computer Technology and Development* 21, 118–120 (2011)
10. Zhang, G., Chen, Y.: Information Semantic Relativity Retrieval Based on Domain Ontology. *Computer Engineering* 37, 34–37 (2011)

Transmission and Distribution Network Analysis Platform Based on FastDB

Zhangzhuo Dong¹, Xiaoting He¹, Jianxin Li², and Kun Li³

¹ Xi'an Shiyou University, Xi'an, China

² Erdos Electric Power Bureau Refresher Test Management Office, Erdos, China

³ Xi'an University of Science and Technology, Xi'an, China

dongzzxa@qq.com

Abstract. It's of great significance to develop an open and expansive distributed analysis platform for power transmission and distribution network. In this paper, a system based on IEC61970-301 common information model and FastDB memory database is introduced. The characteristics of FastDB and the directions for how to use it is illustrated firstly. Then, the map method for storing CIM in FastDB is given. At last, the interface efficiency of FastDB memory database is tested, proving that CIM storage proceeded by memory database is with feasibility and high efficiency.

Keywords: Power Network Analysis Platform, CIM, Memory Database.

1 Introduction

Transmission and distribution network analysis software is one of the important tools in the planning, design, operation and research of the power transmission and distribution system [1~3]. At present, the data model, in all kinds of analysis service software, is only designed for the function and the characteristic of the each application requirements. Proprietary data model had limited the information interaction between the software itself and the third party software; hence the openness and extensibility of software can not be guaranteed. The traditional network analysis software is stand-alone version software, and can not meet the requirements of the current information society. Therefore, it is of great importance to develop a commonly used, opened, networked general analysis software of transmission and distribution grid.

IEC61970-301 Common Information Model standard formulated by International Electrotechnical Commission is to describe the abstract model of main object of power system [4]. It can be used as all kinds of the based model involved with power system analysis software. The transmission and distribution network analysis software based on CIM can achieve good open requirements.

Since it is difficult to manage numerous information and data in computational analysis and a lot of resource maintenance and information cannot be shared in standalone way in transmission and distribution network, the platform uses the

distributed architecture to realize the data centralized management and information shared. Distributed transmission and distribution system analysis platform supports multi-user access at the same time; efficient storage of power network model data is a key technology which is required to resolve in system [5~6].

At present, the memory database is considered as a convenient and efficient information storage technology [7~8]. Since, the FaseDB, one of the memory databases, is used in the Distributed transmission and distribution network analysis software platform to store data [9]. In this paper, firstly summarizes the structure of CIM model and storage mode of CIM model in relational database, analyzes the characteristics and usage method of FastDB memory database, then it gives the storage method of CIM model in memory database FastDB, designs memory data procedure based on CIM model, last tests the efficiency of memory database FastDB and verifies the feasibility and the high efficiency of CIM model storage by using memory database.

2 Common Information Model

CIM is the abstract model which contains all major power system objects, including the public class and attribute of these objects; it can be applied to any software system reference to the power system.

CIM adopts object-oriented pattern to design and unified modeling language to describe. In order to make CIM easy to be designed, understood and browsed, CIM is divided into groups according components related with power system and is defined as a lot of logic packages that each logic package stands for certain part of the power system model. Object may have links across the boundary of multi-package with others, each application will use the information indicated by multi-package.

A package contains one or more class diagrams, and shows all classes and their relationship in the graphics mode, then defines each class in the form of characters according to the kind of attributes of class and the relationship with other class. The connection between classes can be established through three static relationships such as generalization, association and aggregation.

3 Memory Database FastDB

FastDB is an efficient, open source memory database system. The database and the application using this database are kept in the uniform memory space, to access database by memory map mode in the task of application program with no need for task switch and data transmission. Database caching management hardly raises the cost of FastDB, meanwhile, it is not required for FastDB to transmit data in database document and buffer pool. Therefore, FastDB is faster than traditional database.

FastDB supports transactions, online backup and automatic recovery after a system crash. A transaction commit protocol automatically updates database according to a shadow root page algorithm. Online backup and crash recovery provide high availability to critical applications [10].

FastDB makes some optimization in aspect of data querying, carries on query optimization by T-tree index, contrary indicator and query parallelization[11]. Therefore, the query efficiency is highly improved with other database.

The data structure is defined on the basis of table and field in the C# language module provided by FastDB which is similar to the table structure in SQL database. CIM model is object oriented, and makes the class, attribute and relationship as a basic unit. Through the mapping mechanism, a class will be converted to a table, an attribute to a column, a living example to a row and the relationship between two classes will be expressed by the main foreign key or association relationship.

FastDB achieves database table applications in the way of C++ interface. The data in FastDB is stored in table, a table corresponds to a C++ class and a record within the table corresponds to a class instance.

A class corresponding to the table in FastDB defines the table field by defining macro TYPE_DESCRIPTOR(field_list), and defines the class method by defining macro CLASS_DESCRIPTOR(method). field_list is a listing described by class field.

The following kinds of field_list: KEY macro, FIELD macro, SUPERCLASS macro and RELATION macro.

KEY macro formats:KEY (name, index_type)

The name is field name, and index_type is lead field. index_type values HASHED or INDEXED.

If index_type is in the form of HASHED, this filed is a key field and a hash table will be created.

If index_type is in the form of INDEXED, this field is a key field and a T tree will be created.

The macro SUPERCLASS is used to describe the field in the parent class of the class. Relative to the class corresponding to the table, SUPERCLASS(name) specifies the base class of current class.

RELATION (reference, inverse_reference) macro establishes the relationship between one class (table) and another one, reference corresponds to a pointer or pointer array to a class (table). When it refers to the pointer, it is one-to-one relationship between the class and other one, when it refers to the point array, it is one-to-many relationship. inverse_reference is a field pointing to a contrary pointer included in current table, and it values for pointer or pointer array. The relationship described by RELATION may be one-to-one, one-to-many and many-to-many.

OWNER (reference, inverse_reference)

FastDB will automatically update contrary pointer, and use reverse pointer to optimize inquiry efficiency.

The one-to-many or many-to-many relationship is defined between one class of its owe type and another. All specified member records would be deleted when owner record is deleted. The macro RELATION is used to declare when the member record points to owner class.

Following is to define the description methods of Switch class, Breaker class, Reclose Sequence class of CIM 301 protection package in FastDB.

4 Mapping of CIM in FastDB

The implementation of CIM model in FastDB: The direct method is one-to-one mapping and uses the way of defining C++ class to realize the table structure according to the C++ interface provided by FastDB, thus the mapping from class in CIM to table can be realized and the relationship between a class and another one in CIM model can be solved by defining the relationship between classes in FastDB.

Each of the children class and parent class with inheritance relationship can be mapped as a table and also can be mapped as two tables automatically in database by directly using inherited feature of the C++ class in FastDB.

Association relationship and aggregation relationship between classes in CIM can be realized through the macro defined in FastDB:

(1) One-to-one relationship: Build the quote of opposite side in their owe class (table) such as dbReferenceb, RELATION(b, a) in class A and dbReference<A>a, RELATION(a, b) in class B.

(2) One-to-many relationship: dbArray<dbReference b, RELATION(b,a) in class A and dbReference<A>a, RELATION(a, b) in class B.

(3) Many-to-many relationship: dbArray<dbReference b, RELATION(b, a) in class A and dbArray<dbReference<A>a, RELATION(a, b) in class B.

(4) Aggregation relationship: If class A is composed of class B and C, dbArray<dbReferenceb, dbArray<dbReference<C>c, macro OWNER(b, a) and macro OWNER(c, a) are defined in class A, dbArray<dbReference<A>a,macro RELATION(a, b) and macro RELATION(a, c) are defined in class B and class C.

The way of the above has realized the mapping between classes, relationships in CIM and the FastDB database. The multilayer inheritance relationship exists in CIM model with complex relationship between classes. FastDB stability would go wrong when the relationship between classes reaches a certain level in FastDB, resulting in no guarantee of normal reliable operation. Therefore, it can not be realized according to the entire inheritance relationship in CIM, but when to reduce layers, it means that class realization only realize class which needs to be instantiated.

For example: The relationship between EquipmentContainer and Equipment is aggregation relationship in CIM core package, namely, equipment container is composed of several equipments, Bay class inherits from EquipmentContainer class and PowerTransformer class inherits from Equipment class. The object models mentioned above as shown in Fig.1. As can be seen from the object diagram that the relationship between Bay class and PowerTransformer class is aggregation relationship, but there is no description of aggregation relationship between them in CIM model, the aggregation relationship is acquired from relationship between parent classes. Then, if entirely adopting the relationship given by CIM to describe the relationship between power system equipments that resulting in un conformity with reality, thus, the entity classes are adopt to realize the relationship between classes.

Fig.1 shows the relationship diagram described in CIM about several classes.

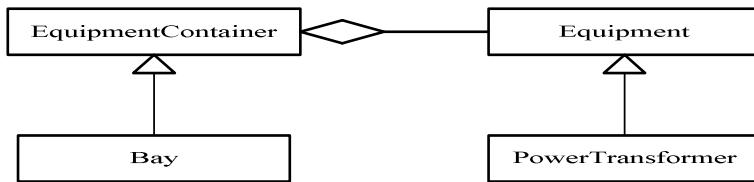


Fig. 1. Class diagram instance in CIM

5 The Design and Implementation of Memory Database Interface

C# interface components provided by memory database can realize the access of the memory database, only the user understands each class function and their relationships, the user can neatly use the interface which is difficult to be used. From the convenience of application considerations, the established CIM model needs one kind of relatively normative forms to design the data access interface.

One of the signs of the open traditional database is adopting sequence query language to access the database, thus, the access to CIM of power grid is built in FastDB adopting SQL language, and the data access in FastDB database should focus on the operation to tables. As long as the realization of table records query, update, add, delete and other operations can meet the requirements.

The following classes are designed for the standard operation interface function.

Regex is a standard regular expression class provided by .NET system framework, being used to pick up information of some fields demanded by user from string.

The strAnalysis class can read the field name, field value, table name and other information based on the given string parameters and rely on Regex class.

FastDBStandardInterface class can provide query, update, add, and delete interface operation function using the corresponding string as the parameters.

Fig.2 shows the classes of the standard operation interface function and the relationship between the classes.

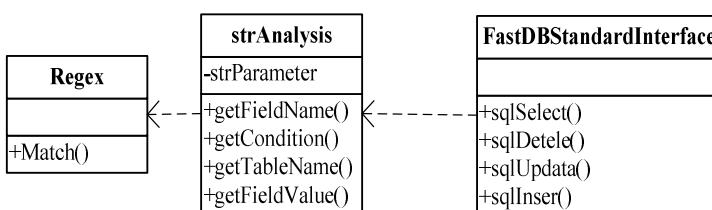


Fig. 2. Class diagram of standard operation interface function

Design the following functions in Regex:

sqlSelect(string sqlQuery):The function is used to query data from memory database.

sqlInsert(string strInsert):The function is used to insert data into memory database.

sqlUpdate(string strUpdate):The function is used to update data of memory database.

sqlDelete(string strDelete):The function is used to delete data specified by database.

The access to FastDB can be realized by the above functions.

Such as visiting the database equipment table Breaker:

The function of sqlUpdate is similar to the following statement: “update Breaker set aliasName='Br1101' where name='1101'”.

Here is a case of creating a table and its field of information in FastDB, and adding data to one as an example to illustrate the cooperative relationship between C# interface classes.

(1) First of all, create a connection with FastDB and establish the object connection of FastDBConnection.

(2) Then the properties of object and function parameters should be set up, including data name, initial size, etc.

(3) The object fields of class FastDbFields is created and is used for setting up the property information including data type, index field, etc.

(4) Object connection creates CreateTable function and adds table and field information to FastDB.

(5) An instance object command of class FastDbCommand is created. The field information is added to object command in the way of the Assign Method of object command and using object fields as parameters.

(6) Insert method of object command is called to insert data value.

(7) Commit method of object connection is called to submit the operation service.

The instance cliEr of class CliError is called to check error messages through the process completion.

Fig.3 shows the process of insert operation by the dynamic sequence diagram of classes:

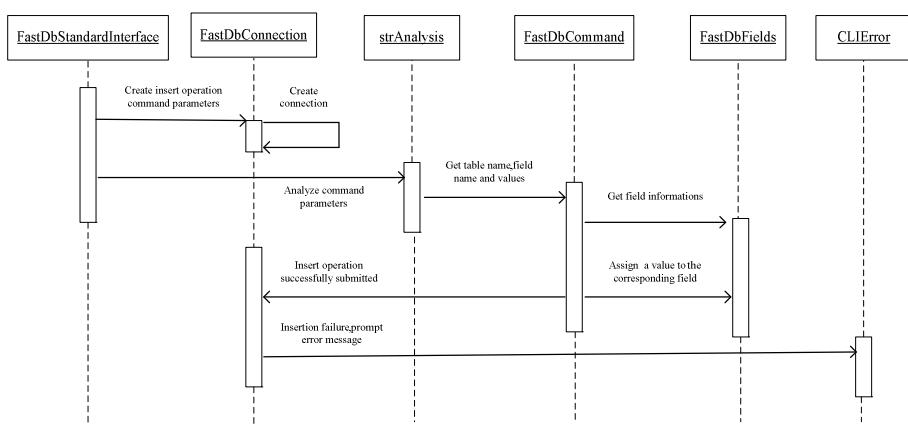


Fig. 3. Sequence diagram of insert operation

6 Interface Testing

In accordance with the above methods, we use C# language develop a FastDB database prototype system of power system can be described by CIM and realize the access to real-time database by the development interface.

The interface testing is conducted at a laptop with a CPU of 1.7GHz Intel Core 2 Duo, 2GB computer memory and operation system of Windows XP.

It exists one-to-one or zero-to-one relationship between class of HydroPump and SynchronousMachine in power production package, so a field synchronous MachineID should be added to the table HydroPump created in SQL2000. Then the table HydroPump and the table SynchronousMachine would be created separately in FastDB. And the main foreign key relationship between the two by means of data dictionary.

(1) Build two tables in the SQL and FastDB by the way of inserting records detailed. The time of inserting 30000 records is showing in table 1.

Table 1. Record chart of testing insert operation

Testing time	Record count	SQL Server		FastDB		Compare
		Time (s)	Speed (record/s)	Time (s)	Speed (record/s)	
1	30000	231.94	129.34	56.07	535.05	The speed of insert
2	30000	234.86	127.74	55.19	543.58	operation in FastDB
3	30000	230.18	130.33	56.86	527.61	is 4.1 times than that
AVG	30000	232.33	129.14	56.04	535.41	in SQL

(2) Proceed delete operation to the two tables in the SQL and FastDB respectively by the way of deleting detailed, the amount of the time for deleting records is showing in table 2.

Table 2. Record chart of testing delete operation

Testing time	Record count	SQL Server		FastDB		Compare
		Time (s)	Speed (record/s)	Time (s)	Speed (record/s)	
1	30000	180.64	166.08	54.68	548.65	The speed of delete
2	30000	176.50	169.97	52.50	571.43	operation in FastDB
3	30000	179.05	167.56	53.72	558.45	is 3.3 times than that
AVG	30000	178.73	167.87	53.63	559.51	in SQL

(3) Proceed update operation to the two tables in the SQL and FastDB respectively by the way of updating detailedly, the amount of the time for updating records is showing in table 3.

Table 3. Record chart of testing update operation

Testing time	Record count	SQL Server		FastDB		Compare
		Time (s)	Speed (record/s)	Time (s)	Speed (record/s)	
1	30000	182.49	164.39	57.26	523.92	The speed of update
2	30000	179.60	167.04	59.43	504.80	operation in FastDB
3	30000	180.53	166.18	59.34	505.56	is 3.0 times than that
AVG	30000	180.88	165.87	58.68	511.43	in SQL

(4) Proceed query operation to the two tables in the SQL and FastDB respectively by the way of querying detailedly, the amount of the time for querying records is showing in table 4.

Table 4. Record chart of testing query operation

Testing time	Record count	SQL Server		FastDB		Compare
		Time (s)	Speed (record/s)	Time (s)	Speed (record/s)	
1	30000	352.14	85.19	168.07	178.50	The speed of
2	30000	368.83	81.34	167.85	178.73	query operation
3	30000	359.49	83.45	158.76	188.96	in FastDB is 2.2
AVG	30000	360.15	83.33	164.89	182.06	times than that in SQL

The testing verifies the correctness and efficiency of query, update, add, delete and other operations operated by the interface to memory database that can satisfy the demand of the system. The testing result shows that the efficiency of real-time database access has improved dramatically than commercial database.

References

1. Wang, H.-F., He, B.-T.: Study on Software of Relay Settings Examination and Rectification. Proceedings of the Electric Power System and Automation 18(2), 90–94 (2006)
2. Wu, W.-C., Sun, H.-B., Zhang, B.-M., et al.: Design of integrated EMS/DTS System Based on IEC61970. Automation of Electric Power Systems 29(4), 53–57 (2005)
3. Kokai, Y.: Recent Development in Open Systems for EMS/SCADA. Electrical Power & Energy Systems 20(2), 111–123 (1998)
4. Draft IEC 61970. Energy Management System Application Program Interface (EMS-API), Part301: Common Information Model (CIM). IEC, USA (2003)
5. Britton, J.P., Devos, N.: CIM-Based Standards and CIM Evolution. IEEE Transactions on Power Systems 20(2), 758–764 (2001, 2005)
6. Yuan, J.-S., Xu, Z.-H., Li, X.-Y.: Research on unified model of substation equipment state information based on CIM. Journal of North China Electric Power University, Vol 36(1) (2009)

7. Zhao, Y., Miao, Y.-F.: On Memory Database and Technology. *Coal Technology* 36(11) (2011)
8. Xue, Y.-F., Wu, J.: Design of Memory Real-time Database Based on CIM. *Computer Engineering and Design* 28(14), 3508–3510 (2007)
9. Kim, S.-W.: Development of a high-performance index manager in a main memory DBMS. *International Journal of Computer Systems Science & Engineering* 18(2), 95–104 (2003)
10. Cao, Y.-L., Ma, W.-X.: Research of implementation scheme for transaction processing in FastDB. *Journal of Minorities Teachers College of Qinghai Teachers University* 21(2) (2011)
11. Zhou, X.-Y., Qin, X.-P., Xu, Z.: A High Efficient Transaction Committing and Recovery Scheme for Parallel Main Memory Database System. *Journal of China University of Mining & Technology* 38(1) (2009)

Subjective Testing System Based on Chinese Word Segmentation

Jiao Cui-Ling¹ and Wang Jian-Ping²

¹ Experiment Center, Henan Institute of Science and Technology, 453003, XinXiang, China

² School of Information Engineer, Henan Institute of Science and Technology,
453003, XinXiang, China

{jiaocuiling1969,xunji2002}@163.com

Abstract. Subjective testing system based on computer can improve the efficiency of marking the papers greatly. However, the subjective testing system which is used widely at present can only solve the scoring process of objective test which have given answers. Because of the flexibility of objective test's answers, researching the scoring of objective test becomes the problem needed to be solved in the examination system. In this paper, we put forward a subjective testing system based on Chinese word segmentation. We described the introduction, requirement analysis and the core modular of the system in detail. At last, we do the software testing, due to the test results, we know that, it is feasible and reliably by using Chinese word segmentation to realize the subjective testing system.

Keywords: Subjective Testing System, Chinese Word Segmentation, Software Testing.

1 Introduction

Examination is an important link of the current system in our country and all kinds of review selection. Text test is divided into objective test and subjective test. Scoring objective test can use computer, but subjective test can't use computer to get effective judgment. So it has the necessary to research how to score the subjective test and decrease the workload of teachers.

Now academics at home and abroad have started to research this area, and have got some achievements. There are some application systems abroad, such as Project Essay Grade, Latent Semantic Analysis, Educational Testing Service and so on. At home, there are systems used for computer text-entry and C programming language, and some research methods as concept network have been put forward. But some of these achievements and methods have strong field limitations. Some can't score the specific scores. In a word, there is not method and system can be used to scoring the similarity of subjective test's text semantic limited in Chinese field.

In Chinese, most language processes start from word segmentation and part-of-speech tagging. These two steps tokenize a sequence of characters without delimiters

into words and predict a syntactic label for each segmented word. Chinese word segmentation involves mainly two research issues: word boundary disambiguation and unknown word identification. In most of the current systems, these are considered to be two separate tasks and are dealt with using different components in a cascaded or consecutive manner. This paper, we put forward a subjective testing system based on Chinese word segmentation. We described the introduction, requirement analysis and the core modular of the system in detail.

2 Requirement Analysis

2.1 The Description of System Function

The design of Subjective testing system uses the objective-oriented programming technology. Subjective testing system is faced to system administrators, teachers and students. In this system, system administrators can login the system, add users and delete users.

For teachers, online examination system is like a question platform. Teachers can login the system, give titles and check grades.

For students, online system is like a bulletin board to publish grades. Students can login the system, look at grades and do exercise.

2.2 Function Module Division

Subjective testing system has functions that it can add and delete users, check grades, take tests for students, give titles by teachers etc. According to the requirements analysis, the system can be divided into five modular.

(1) Login Modular

The modular can enter into different pages according to identities of different users. Teachers and students can enter into different pages after logging.

(2) User Management Modular

It can provide all the information of users in this modular. Administrator can add and delete users.

(3) Teachers' Homepage Modular

This modular shows all information of grades to teachers. Teachers can select different subjects to look at the students' grades. Besides it contains the entrance of question modular.

(4) Question Modular

There are two pages in this modular. One is setting up new a test paper, other is add question to test papers.

(5) Students' Homepage Modular

The modular contains all grades of candidates and there are all test papers that have been prepared. Clicking "go into the examination room" that the test paper which is

need to be answered corresponds to can log into the test page. Clicking "details of grades" those candidates' grades corresponds to can enter into the details of grades modular.

(6)The Details of Grades Modular

It shows the detailed list of student's grades.

(7)Examination Modular

In this modular, it will show contents of test paper to students after clicking "start the test". When click "hand in the paper", the page will be logged off and exit. Now the function of checking papers has been accomplished, and grades have been recorded. When students log in the system next, they will look at the grades.

The function modular of Subjective testing system shows as Fig. 1.

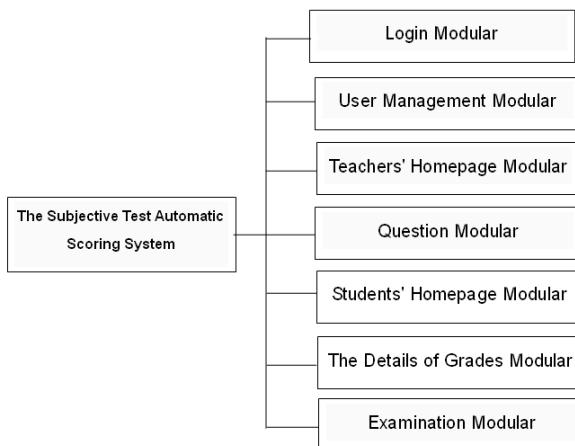


Fig. 1. The function modular of Subjective testing system

3 Core Modular of the System

Scoring function modular is the core modular of the system. It involves the theoretical basis of scoring the subjective test, and writes algorithms based on this. Thereby the function of automatic scoring is achieved.

3.1 The Theory of Statements Similarity

The first is to calculate the similarity of statements, there are a variety of methods of sentence similarity calculation, this paper adopts calculation methods based on the statements of the surface features similarity, describing in detail as follows:

If there are two sentences A, B, the similarity of sentences A and B can express with formula 1:

$$SentenceSimilarity(A, B) = 2 * \frac{SameWords(A, B)}{Number(A) + Number(B)} \quad (1)$$

Among it, $SentenceSimilarity(A, B)$ means the similarity of the sentences A and B, $SameWords(A, B)$ means the number of same words in the sentences A and B, $Number(A)$ and $Number(B)$ means the number of words in the sentences A and B respectively.

3.2 The Scoring Algorithm Based on Chinese Word Segmentation

In the process of scores firstly, after filtering the punctuation of the student's answer and the standard answer respectively, call the divided word function and return the array words and the number of words. Second, compare the two words array after filtering the repeat words out, record the same number of words, and pray for the similarity of statements. Again, searching the key word in points words and student's answer in turns, if successful, record the key word and its range. Finally, according to the formula 2 calculate student's total points of this question, then, add up to the total score of test paper.

$$totalscore = score + (questionscore - keynum) * similary \quad (2)$$

Basic words segmentation algorithm are described below:

- (1) If the points of words stu's length is greater than 0, execute 2;Otherwise return the number words I, exit function.
- (2) If the length of stu is equal or greater than t, execute 3;Otherwise, execute 8.
- (3) Make MaxLength = t.
- (4) If MaxLength! = 1, execute 5; otherwise do 7.
- (5) Intercept reversely the string whose length is MaxLength from stu and assign it to string str,search in the words library, If f successful, it is considered a word or words whose length is MaxLength and stored in the string array std [] and update it, turned to 4;otherwise;execute 6.
- (6) Make MaxLength subtract 1 and give it to MaxLength, turn to 4
- (7) If MaxLength == 1, intercept reversely one word of the storage whose length is 1 from stu and update stu, turn to 2.
- (8) If stu's length is less than t, we make t equal to stu's length, then turn to 1. The specific algorithm refer to the algorithm realizing.

4 Conclusion

In order to test the system based on Chinese Word Segmentation, we give the relative questions, the reference answers, the scoring rules, and the total score. Table 1 shows the scoring rules of the questions.

Table 1. Scoring Rules Of The Questions

Questions(In Chinese)	The reference answers(In Chinese)	scoring rules(In Chinese)	Total score
“三个代表”指的是什么？	“中国共产党始终代表中国先进生产力的发展要求,中国共产党始终代表中国先进文化的前进方向,中国共产党始终代表中国最广大人民的根本利益。”	“先进生产力的发展要求”-5 “先进文化的前进方向”-5 “最广大人民的根本利益”-5	15

The scoring rules can be used by the system according to Chinese word segmentation, and the testing results can be seen as table 2. In this table, we list the relative answers for the four students, and give the relative scores in details by using Chinese word segmentation.

Table 2. Testing results of four students

Student number	Answers of students	Relative scores in details using	The scores of students	total of students
1	"中国共产党始终代表生产力的发展要求,始终代表人民的根本利益."	"生产力的发展要求"-3 "人民的根本利益"-3	6	
2	"中国共产党始终代表人民的根本利益,始终代表先进生产力的发展要求."	"先进生产力的发展要求"-5 "人民的根本利益"-3.	8	
3	"代表中国最广大人民的根本利益,代表中国先进生产力的发展要求."	"中国最广大人民的根本利益"-5 "中国先进生产力的发展要求"-5	10	
4	"代表中国先进文化的前进方向,代表先进生产力的发展要求,代表中国最广大人民的根本利益."	"中国先进文化的前进方向"-5 "先进生产力的发展要求"-5 "中国最广大人民的根本利益"-5	15	

From the five answers above, we can draw conclusions as follows:

- (1) Exchanging statement order will not affect the scores.
- (2) The more similar the answer statements of the same key word are, the higher the score is.
- (3) The more complete the answer is, the higher the score is.
- (4) As long as the words appears in student's answer are the same as the words in standard answer, the student can get full marks.

When the same words appearing in key words, it is possible to affect keyword scores, if the words appear in a student's answer cover the words in the standard answer and much more, resulting in the met degrees is low thus affecting scores.

References

1. Sun, T.-L., Liu, Y.-J., Yang, L.-H.: An ambiguity discovery algorithm on Chinese word segmentation based dictionary. In: Proceedings of the 2009 2nd Pacific-Asia Conference on Web Mining and Web-Based Application, pp. 39–42 (2009)
2. Zhu, C.-H., Zhao, T.-J., Zheng, D.-Q.: Joint Chinese word segmentation and POS tagging system with undirected graphical models. *Journal of Electronics and Information Technology*, 700–704 (2010)
3. Jia, H., Lin, C.: Chinese word segmentation based on the improved particle swarm optimization neural networks. In: International Conference on Cybernetics and Intelligent Systems, pp. 1024–1027 (2008)
4. Yi, Z.-A., Lv, J.: Application of Chinese word segmentation based on linguistic environment analysis in text information filtering system. In: International Conference on Electronic Computer Technology, pp. 467–470 (2009)
5. Fu, G., Kit, C., Webster Jonathan, J.: Chinese word segmentation as morpheme-based lexical chunking. *Information Sciences*, 2282–2296 (2008)
6. Zhang, R., Keiji, Y., Eiichiro, S.: Chinese word segmentation and statistical machine translation. *ACM Transactions on Speech and Language Processing*, 1074–1078 (2008)
7. Zhou, J., Jiang, G., Mao, X.: Subjective quality analyses of stereoscopic images in 3DTV system. *IEEE Visual Communications and Image Processing*, 1040–1048 (2011)
8. Rouse, D.M.P., Romuald, Le Callet, P.: Tradeoffs in subjective testing methods for image and video quality assessment. *International Society for Optical Engineering*, 2070–2074 (2010)
9. Gote, N., Tuomas, L., Lindroos, P.: Measuring multivariate subjective image quality for still and video cameras and image processing system components. *The International Society for Optical Engineering*, 456–464 (2008)
10. Kaoru, A., Kohei, N.: A nonlinear image processing system for beautifying face image using interactive evolutionary computing and its subjective test. *IEEJ Transactions on Electronics, Information and Systems*, 576–583 (2011)
11. Peter, W.: Subjective impression to objective product noise testing. *Acoustics Bulletin*, 33–35 (2008)

An Efficient Algorithm for Reconstruction of Discrete Sets from Horizontal Projections with Absorption

Chen Junyu¹, Gao Aili², and Zhang Chuanlin^{1,2,*}

^{1,2} College of Information Science and Technology,
Jinan University, 510632, Guangzhou, China
time240326147@126.com

Abstract. This paper considered that two projections along the left and right horizontal directions uniquely determine a binary matrix when the absorbed coefficient is special. For the weakness of computational complexity of projection difference, an improved algorithm is proposed to reconstruct binary matrices along the diagonal projections based on determining conditions of sequence consistency. Furthermore, comparing with the existing algorithm, it's more efficient to search solutions.

Keywords: Discrete tomography, Discrete model with absorption, Reconstruction algorithm.

1 Introduction

The main problem of Discrete Tomography (DT) in [1] is to recover some geometrical properties of an object considered as a discrete set of Z^d , with $d \geq 2$, from partial and sometimes inaccurate information. In this paper we deal with the specific problem of reconstructing a planar emitting body from its horizontal projections in the case that both the body and its surrounding material absorb part of the emitted radiations (Emission DT). More precisely, if I_0 is the initial intensity of the emission then the intensity I that can be detected at distance x follows from the law:

$$I = I_0 \cdot e^{-\mu x} \quad (1)$$

where μ is the absorption around the emitting point. We simplify the model by considering the absorption function to be constant $\mu_0 = \ln(\frac{1+\sqrt{5}}{2})$. Finally, we set $\beta = e^{\mu_0}$. As $\mu \geq 0$ and have $\beta \geq 1$.

In section 2, we introduce some definitions and notations of words. In section 3, we provide an efficient algorithm. In section 4, we apply it to binary matrices

* Corresponding author.

reconstruction from diagonal projections. Furthermore, compare it with the existing algorithm.

2 Preliminaries

Let our environment be the integer lattice $Z \times Z$ and F be a finite set of points. The minimum (discrete) rectangle containing F , and having dimensions $m \times n$, can be naturally represented by a binary matrix A , $a_{i,j} = 1$ if and only if the lattice point $(i, j) \in F$. The vectors of the left and right horizontal absorbed projections of A , say L_β and R_β , respectively, are defined as follow:

$$L_\beta = (l_1, l_2, \dots, l_m), \text{ where } l_i = \sum_{j=1}^n a_{i,j} \beta^{-j};$$

$$R_\beta = (r_1, r_2, \dots, r_n), \text{ where } r_i = \sum_{j=1}^m a_{i,j} \beta^{-n+j-1};$$

for each $1 \leq i \leq m$. The coefficient $\beta \geq 1, \beta \in R$, represents the absorption of the material. A binary matrix A is called a realization of L_β and R_β if A has L_β and R_β as left and right absorbed projections.

Example 1. The left and right projections are

$$L_\beta = (\beta^{-2} + \beta^{-4}, \beta^{-3} + \beta^{-4}, \beta^{-4}, \beta^{-1} + \beta^{-5}),$$

$$R_\beta = (\beta^{-2} + \beta^{-4}, \beta^{-2} + \beta^{-3}, \beta^{-2}, \beta^{-1} + \beta^{-5}),$$

$$\text{Where } A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

is a realization of L_β and R_β .

The simple notions defined above allow us to translate the classical problem of the reconstruction of a discrete planar set from its projections into our framework.

$$\text{Reconstruction DA2D}(L, R, \beta_0)$$

Input: A real coefficient $\beta \geq 1$, two integers m and n , and two vectors $L_\beta, R_\beta \in R^m$.

Task: Construct a binary matrix A of size $m \times n$ which it's a realization of $L_\beta, R_\beta \in R^m$.

In [2,3,4,5], authors focus on the case $\beta = \frac{1+\sqrt{5}}{2} = \beta_0$, the well known golden ratio, a number often encountered in several combinatorial and geometrical problems. In particular, the property

$$\beta^{-i} = \beta^{-i-1} + \beta^{-i-2} \quad (2)$$

makes this coefficient very interesting from a mathematical point of view.

Firstly, we introduce some definitions and notations of words in [6].

2.1 β_0 -representation

A binary word $w = a_1a_2\dots a_n$, with $n \in N$, is called a finite n-length representation in base of β_0 (briefly β_0 -representation) of $r \in R$ if $r = a_1\beta_0^{-1} + a_2\beta_0^{-2} + \dots + a_n\beta_0^{-n}$. The β_0 -representation is said to be a β_0 -expansion if it doesn't contain a 011 subword.

Naturally, each number r can have several n-length β_0 -representation, but the definition guarantees the existence of a single n-length β_0 -expansion.

Example 2. Using Eq.(2) it is easy to check that the following words are 11-length β_0 -representation of the same number:

$$\begin{aligned} w_1 &= 01011010000 \\ w_2 &= 01011001100 \\ w_3 &= 01100001100 \\ w_4 &= 10000010000 \end{aligned}$$

but only the last one is β_0 -expansion.

2.2 1D-switchings

According to the definitions in [7], we call elementary 1D-switch the substitution, inside a β_0 -representation, of an occurrence of the subword 100 with the subword 011. Let x_k be a generic digit, with $k \geq 0$. We call 1D-switch the substitution of an occurrence of the subword $1(0x_k)^k 00$ with $0(1x_k)^k 11$. Each switch can be obtained by one or more consecutive applications of elementary switches, its application doesn't change the value of β_0 -representation (see Example 2).

2.3 Uniqueness

We start by recalling the following theorem, proved in [5], which characterizes 1D-switching components when the absorption is β_0 .

Theorem 1. Let $w = a_1 a_2 \dots a_n$ and $w_1 = b_1 b_2 \dots b_n$ be two different β_0 -representation of the same number. If $\sum_{i=1}^n a_i \beta_0^{-n+i-1} = \sum_{i=1}^n b_i \beta_0^{-n+i-1}$, then $a_i = b_i$, for all $1 \leq i \leq n$.

Corollary 2. A binary matrix A is uniquely determined by the vectors of its left and right horizontal absorbed projections.

3 An Efficient Algorithm

In [8,9,10,11,12], authors propose several algorithms to solve the problem of binary matrices reconstruction, based on them, we propose the efficient algorithm. In this section, L_{β_0} and R_{β_0} is initialized. If the binary sequences $a_{i,j} a_{i,j+1} \dots a_{i,j+n}$ which compatible with L_{β_0} and R_{β_0} exist, the solution is unique, false otherwise. An efficient algorithm is as follow:

3.1 The Function ROW

For $x = a_1 \beta_0^{-1} + a_2 \beta_0^{-2} + \dots + a_n \beta_0^{-n}$, the length is initial, then the coefficients of $\beta_0^{-1}, \beta_0^{-2} \dots \beta_0^{-n}$ can be changed into 1 by switching some $1(0x_k)^k 00$ into $0(1x_k)^k 11$ and the corresponding sequence $b_1 b_2 \dots b_n$ is obtained. If the length exceeds the initial one, the solution doesn't exist. In this function, we use hash in language Perl.

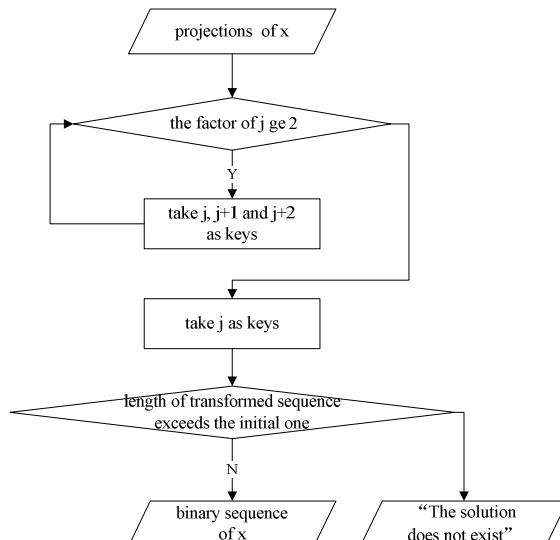


Fig. 1. The function ROW

3.2 The Function TRAN

Search the first position of 00

For a binary sequence $b_1 b_2 \dots b_n$, we search the first position of 00 in the initial sequence on the left, then mark its previous position as k .

Determine the position of switch

The algorithm starts at the end of the right, search for the position of switch 100, where a switch $100 \rightarrow 011$ increases the value of R_β . For each row of a binary matrix, we should search the largest position j which meet the condition as follow:

$$a_{i,j} = 1, a_{i,j+1} = 0, a_{i,j+2} = 0 \quad (3)$$

Perform the switch

If a switch $100 \rightarrow 011$ is performed at positions $j, j+2, \dots, j+2k$, the next one can be performed at position $j+2k+2$ or at a position on the left of $j-1$.

End the circulation

In the above steps, we find the position of $100 \rightarrow 011$ and perform it. In other words, the position j moves forward constantly until the position j move to k , as the left sequence doesn't perform the switch, then the algorithm end the circulation. Finally, it transforms a maximal $1(00)^k 00$ sequence into a $0(1x_k)^k 11$ sequence.

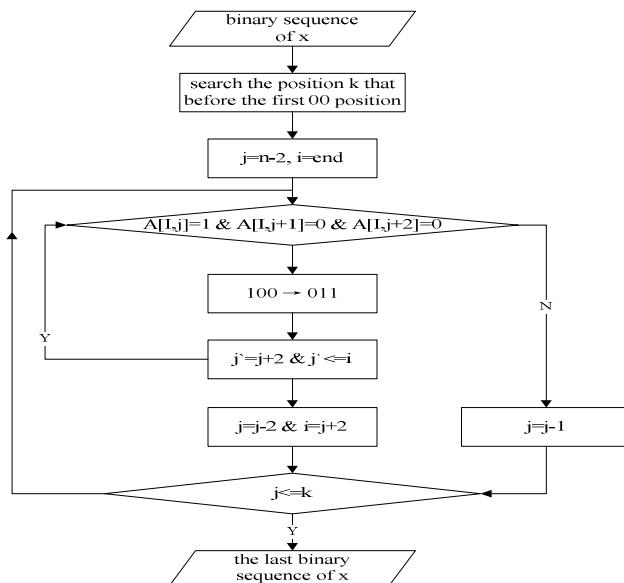


Fig. 2. The function TRAN

3.3 Output Conditions

The previous algorithm uses the difference of the right projections to decide the existence of the sequences along the left and right projections, while our algorithm uses the consistency of the sequence. Through the analysis, it may appear four situations, where z is the sequence after the sequence x performs the switches, z' is the negative sequence of x , z'' is the negative sequence of z , y' is the sequence after y performs switches.

- (i) The left projection x and right projection y don't transform, then $z' = y$
- (ii) The left projection x transforms and the right projection y don't transform, then $z'' = y$

- (iii) The left projection x and right projection y don't transform, then $z'' = y'$

Through the three conditional judgments above, it may also appear two sequences aren't equal, but the sequences of reconstruction still exist, so we add a determining condition(iv).

(iv) After (iii) we obtain the sequence z'' , so we execute the function TRAN again to obtain z''' sequence, then $z''' = y'$.

Example 3. Let $l = \beta_0^{-4}$, $r = \beta_0^{-1}$, $\text{length} = 7$, then we have

$$x = \text{ROW}(l, 7) = 0001000, z = \text{TRAN}(x) = 0000110; z'' = 0110000;$$

$$y = \text{ROW}(r, 7) = 1000000, y' = \text{TRAN}(y) = 0101011; z'' \neq y',$$

$$z''' = \text{TRAN}(z'') = 0101011, \text{then } z''' = y'.$$

Finally, the sequence z is the solution.

3.4 The Flow of the Algorithm

Step1: Input $L = (x_1, \dots, x_m), R = (y_1, \dots, y_m)$, where

$$L = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} \beta_0^{-1} \\ \beta_0^{-2} \\ \vdots \\ \beta_0^{-n} \end{pmatrix}, R = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} \begin{pmatrix} \beta_0^{-1} \\ \beta_0^{-2} \\ \vdots \\ \beta_0^{-n} \end{pmatrix}$$

Step2: Invoke $\text{ROW}(x_i, m), \text{ROW}(y_i, m)$, then transform $x_i, y_i (1 \leq i \leq n)$ into binary sequences.

Step3: Judge the determining conditions are met or not, invoking the function TRAN. If one of the determining conditions is met, output the sequence, else output the solution doesn't exist.

4 Empirical Analysis

We have implemented our algorithms in language Perl, using Linux operating system. The experiments were run on Pentium(R) Dual-Core CPU with 2G of memory.

4.1 Examples of Binary Matrices Reconstruction

Example 4. The left projection and right projection vector are given, and the size is 7×7 . In the following, we describe step by step effects of the call RECROW (L,R) on every row vectors.

$$\begin{aligned} L &= (2\beta_0^{-1} + \beta_0^{-3}, \beta_0^{-1} + 2\beta_0^{-3} + \beta_0^{-7}, \beta_0^{-1} + \beta_0^{-2}, \\ &\quad \beta_0^{-1} + \beta_0^{-2} + \beta_0^{-5}, \beta_0^{-1} + \beta_0^{-6}, \beta_0^{-1}, \beta_0^{-4}); \\ R &= (2\beta_0^{-1} + \beta_0^{-5}, \beta_0^{-1} + 2\beta_0^{-3} + \beta_0^{-7}, \beta_0^{-1} + \beta_0^{-5} + \beta_0^{-7}, \\ &\quad 2\beta_0^{-1} + \beta_0^{-7}, 2\beta_0^{-2} + \beta_0^{-6}, \beta_0^{-1} + \beta_0^{-2} + \beta_0^{-4} + \beta_0^{-6}, \beta_0^{-1}). \end{aligned}$$

Take the fifth row for an example, $l_5 = \beta_0^{-1} + \beta_0^{-6}, r_5 = 2\beta_0^{-2} + \beta_0^{-6}$.

Step1: Execute $ROW(l_5, 7)$, we obtain the initial sequence 1000010 of x . In similar way, execute $ROW(r_5, 7)$, the initial sequence 0111010 of y can also be obtained.

Step2: Execute $TRAN(x)$, the position $k = 1$;

Step3: From the fifth position on, check whether (3) is met or not. It isn't met, 5-1=4.

Step4: Repeat step3, we have it's met in the 1st position, then execute $TRAN(x)$, 0101110 can be obtained.

Step5: Finally, as (ii) is met, so we output 0101110.

The matrix is reconstructed and it's showed in Fig.3, where a black pixel represents 1 and a white pixel represents 0, the same as follow.

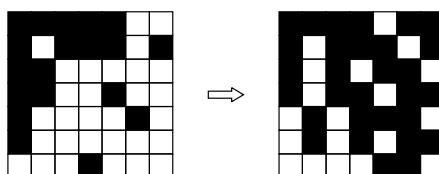


Fig. 3. The initial matrix from the two projections and the reconstructed matrix

Example 5. The left and right projections along the diagonal direction are given, and the size is 7×7 .

$$L = (0, 0, \beta_0^{-1}, \beta_0^{-1} + \beta_0^{-4}, \beta_0^{-2} + \beta_0^{-5}, \beta_0^{-2} + \beta_0^{-5}, \beta_0^{-2} + \beta_0^{-5} + \beta_0^{-6}, \\ \beta_0^{-2} + \beta_0^{-5}, \beta_0^{-2}, \beta_0^{-2} + \beta_0^{-3}, \beta_0^{-1}, \beta_0^{-1} + \beta_0^{-2}, \beta_0^{-1});$$

$$R = (0, 0, \beta_0^{-1} + \beta_0^{-2}, \beta_0^{-1} + \beta_0^{-2} + \beta_0^{-3}, \beta_0^{-1} + \beta_0^{-2} + \beta_0^{-3}, \\ \beta_0^{-2} + \beta_0^{-3} + \beta_0^{-4}, \beta_0^{-2} + \beta_0^{-3} + \beta_0^{-4} \beta_0^{-5}, \beta_0^{-2} + \beta_0^{-3} + \beta_0^{-4}, \\ \beta_0^{-2} + \beta_0^{-3}, \beta_0^{-2} + \beta_0^{-3}, \beta_0^{-1} + \beta_0^{-2}, \beta_0^{-1} + \beta_0^{-2}, \beta_0^{-1});.$$

Take the 7th row for an example,
 $l_7 = \beta_0^{-2} + \beta_0^{-5} + \beta_0^{-6}$, $r_7 = \beta_0^{-2} + \beta_0^{-3} + \beta_0^{-4} + \beta_0^{-5}$.

Step1: $ROW(l_7, 7) = 0100110$, $ROW(r_7, 7) = 0111100$;

Step2: $ExecuteTRAN(l_7) = 0011110$;

Step3: As the determining condition (ii) is met, $TRAN(l_7) = ROW(l_7, 7)$, output the sequence 0111100 .

The matrix is reconstructed from two projections along the diagonal direction and it's showed in Fig. 4.

We note that the black pixel of the initial matrix is scattered, while it's centralized in the final one. It's important to image reconstruction.

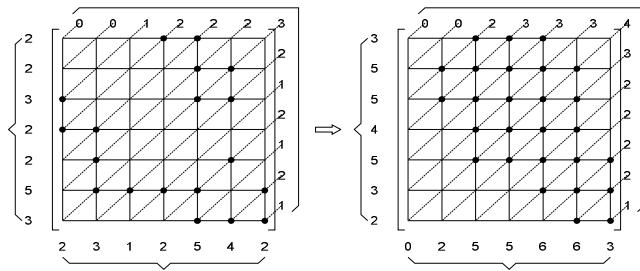


Fig. 4. The initial matrix and the reconstructed matrix

4.2 The Comparison of Execution Time

In the experimental phase we tested two algorithms to search the unique solution, table 1 is the correlation time table of two algorithms. Data shows that the efficient algorithm we propose costs less time. Furthermore, no matter the CPU or the kernel running time, it displays a clear superiority. The results are summarized in table 1. PA and EA represent previous algorithm and efficient algorithm respectively. The column Real contains the real time of two algorithms used on different size of binary matrices the same as column User and Sys. The unit of time is minute and second.

Table 1. Time table of two different algorithms

	Real Time PA	EA
100*100	0m0.735s	0m0.279s
200*200	0m3.970s	0m0.507s
300*300	0m12.753s	0m1.093s
400*450	0m37.513s	0m1.748s
500*500	0m57.860s	0m2.538s
550*600	1m29.696s	0m3.507s
700*650	2m18.431s	0m4.833s
800*800	4m1.632s	0m6.460s
900*900	5m38.907s	0m7.778s
1000*1000	7m36.787s	0m9.828s
1000*1500	13m9.886s	0m15.280s
2000*2000	48m37.902s	0m39.754s
2500*2500	96m46.763s	1m3.315s
3000*4000	296m43.294s	2m2.590s
5000*5000	769m12.919s	4m48.469s
6000*5500	1149m17.025s	6m9.149s
7000*7500	2626m38.142s	9m53.371s
8000*8000	3405m12.114s	11m45.621s
8500*9000	4926m15.660s	15m34.416s
10000*10000	8268m40.067s	20m49.404s

User Time PA	EA	Sys Time PA	EA
0m0.548s	0m0.156s	0m0.090s	0m0.010s
0m3.876s	0m0.377s	0m0.013s	0m0.007s
0m12.547s	0m0.871s	0m0.064s	0m0.021s
0m37.238s	0m1.606s	0m0.030s	0m0.019s
0m57.524s	0m2.284s	0m0.021s	0m0.031s
1m29.345s	0m3.150s	0m0.038s	0m0.029s
2m18.053s	0m4.271s	0m0.058s	0m0.044s
4m0.507s	0m5.923s	0m0.080s	0m0.038s
5m35.266s	0m7.537s	0m0.124s	0m0.025s
7m34.566s	0m9.280s	0m0.122s	0m0.074s
13m5.084s	0m14.811s	0m4.641s	0m0.121s
48m24.906s	0m39.402s	0m12.553s	0m0.133s
96m28.805s	1m2.301s	0m17.193s	0m0.315s
294m57.558s	2m2.078s	1m43.913s	0m0.212s
764m40.690s	4m47.356s	4m27.597s	0m0.564s
1144m54.998s	6m8.272s	4m14.104s	0m0.574s
2617m37.838s	9m51.843s	8m29.333s	0m0.983s
3395m22.187s	11m44.289s	9m18.554s	0m1.051s
4904m58.138s	15m30.785s	20m30.269s	0m3.049s
8215m10.001s	20m47.136s	17m20.099s	0m1.817s

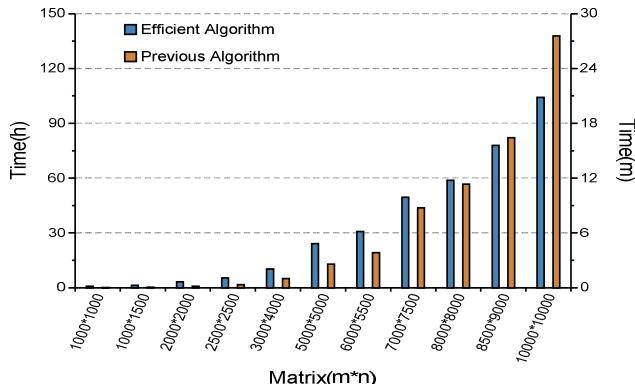


Fig. 5. Real Time of Two different algorithm

As we see from Fig. 5, double y axis bar chart shows that the level of efficient algorithm execution time is minute, while the level of previous one is hour. Comparing to the previous one, it's true that the efficient algorithm using sequence consistency as judgment has feasibility and high efficiency.

5 Conclusion

In this paper we showed an efficient algorithm for reconstructing binary matrices from two horizontal projections. We use the EDT physics model when the absorption coefficient is $\beta_0 = \frac{1+\sqrt{5}}{2}$, then based on the previous algorithm, introducing binary sequences and build function RECROW to realize the reconstruction of binary matrices. Finally, we implemented our algorithm in Linux operation system.

The experimental results show that the efficient algorithm not only can solve the problem of binary matrices reconstruction along horizontal projections, but also can search the solution along the diagonal projections. On some test large matrices, our algorithm speeds up the search for unique solutions.

Acknowledgments. This work is supported by NSF of China (61070165) and Product Research and Development Foundation of Education Ministry and Guangdong Province under Grant (2011B090400458).

References

1. Herman, G., Kuba, A.: *Discrete tomography: foundations, algorithms and applications*. Birkhäuser, Boston (1999)
2. Frosini, S., Rinaldi, A., Kuba: An efficient algorithm for reconstructing binary matrices from horizontal and vertical absorbed projections. *Electronic Notes in Discrete Mathematics* 20, 347–363 (2005)

3. Kuba, Nivat, M.: A sufficient condition for non-uniqueness in binary tomography with absorption. *Theoretical Computer Science* 346, 335–357 (2005)
4. Kuba, Nagy, A., Balogh, E.: Reconstruction of hv-convex binary matrices from their absorbed projections. *Discrete Applied Mathematics* 139, 137–148 (2004)
5. Barcucci, E., Frosini, A., Rinaldi, S.: An algorithm for the reconstruction of discrete sets from two projections in presence of absorption. *Discrete Applied Mathematics* 151, 21–35 (2005)
6. Woeginger, G.W.: The reconstruction of polynomial from their orthogonal projections. *Inform. Process. Lett.* 77, 225–232 (2001)
7. Kuba, A., Nivat, M.: Reconstruction of Discrete Sets with Absorption. In: Nyström, I., Sanniti di Baja, G., Borgefors, G. (eds.) *DGCI 2000. LNCS*, vol. 1953, pp. 137–148. Springer, Heidelberg (2000)
8. Miklos, P., Tech, S.: Discrete tomographic reconstruction of binary matrices using tabu search and classic ryser algorithm. In: *9th International Symposium on Intelligent Systems and Informatics (SISY 2011)*, pp. 387–390 (2011)
9. Miklos, P., Csongor, G.: Discrete tomographic reconstruction of binary matrices using branch and bound method. In: *7th Intellithent Systems and Informatics (SISY 2009)*, pp. 85–88 (2009)
10. Jarray, F., Tlig, C.: A simulated annealing for reconstructing hv-convex binary matrices. *Electronic Notes in Discrete Mathematics* 36, 447–454 (2010)
11. Balazs, P.: A benchmark set for the reconstruction of hv-convex discrete sets. *Discrete Applied Mathematics*, 3447–3456 (2009)
12. Varga, L., Balazs, P., Nagy, A.: Direction-dependency of binary tomographic reconstruction algorithms. *Graphical Models*, 365–375 (2011)

Center-Based Iteration Algorithm of Pre-extracting Support Vectors

Yun Shiwei and Shu Yunxing^{*}

Luoyang Institute of Science and Technology, Luoyang, Henan Province, 471023, China
Lysyx6202@sina.com

Abstract. On the basis of research on the mechanism of the support vector machine (SVM) and the distribution characteristics of support vectors, a class center-based algorithm of pre-extracting support vectors was proposed. First, a hyperplane band was constructed via the class centers of two classes of samples and the width of the band was reduced via iteration; then, the samples within the band were employed in training as an alternative to the entire training samples, thus reducing the training samples and meanwhile excluding some outliers; and finally, the algorithm proposed was tested via artificial data and UCI data, and the results of the simulation experiments indicate that when the influence on the classification precision is modest, the proposed algorithm can significantly increase the training speed.

Keywords: Support Vector Machine (SVM), Class Centers, Pre-Extracting Support Vectors, Iteration.

1 Introduction

The support vector machine (SVM) is a new machine learning algorithm proposed by Vladimir Naumovich Vapnik in the 1990s [1]. This algorithm is established on the theoretical basis of structural risk minimization (SRM), therefore it has such merits as small sample size requirements and strong generalizability, and it can avoid the data disaster and overfitting problems, thus it is widely applied in the field of pattern recognition and so on. In theory, the SVM algorithm is actually to solve a quadratic planning problem under a convex constraint. Consequently, when the quantity of the training samples is too much, the training time will be greatly lengthened, which constitutes a major problem constraining practical applications of the SVM algorithm. In view of this problem, many scholars both at home and abroad have proposed some improved algorithms, which can generally be classified into two categories: one is improved on the basis of optimization algorithms, with typical examples as the block algorithm, the SMO algorithm, the parallel algorithm, and the ensemble algorithm, etc.[2-6]; the other refers to those pre-extracting algorithms based on support vectors. Because the classifying hyperplane of the SVM is determined by few support vectors, therefore if some samples can be pre-extracted before training according to some certain characteristics of the samples, making these pre-extracted samples contain

^{*} Corresponding author.

most of the support vectors and in the meantime effectively excluding some outliers, then the training speed will be consequently enhanced by using these samples to substitute for the entire set of training samples. Currently, the proposed methods include the fuzzy pre-extracting method [7], the center distance ratio method[8], the vector projection-based method [9], fuzzy support vector machine method based on border vector extraction [10] etc.

On the basis of research on the mechanism of the SVM and its geometrical characteristics, a class center-based iteration algorithm of pre-extracting support vectors was proposed. First, a hyperplane band was constructed via the class centers of the two classes of samples and the width of the band was reduced via iteration; then, the samples within the band were employed in training as an alternative to the entire training samples, thus reducing the training samples and meanwhile excluding some outliers; and finally, the validity of the algorithm proposed was verified by means of experiments.

2 The Rationale of the SVM

2.1 Linear SVM

For a dichotomous problem, given a set of samples $S = \{(x_i, y_i) | i = 1, 2, \dots, l\}$, $x_i \in R^n$, $y_i \in \{-1, 1\}$, the SVM is to seek for an optimal hyperplane, and a compromise will be made between the classification interval maximization and the classification error minimization, that is, to minimize the structural risks. If reflected in a mathematical expression, the SVM is to solve the following quadratic planning problem [1]:

$$\begin{aligned} \min \quad & \Phi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \left(\sum_{i=1}^l \xi_i \right) \\ \text{s.t.} \quad & y_i (\omega \cdot x_i - b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned}$$

where C is the compromise parameter, which is used to make a compromise between the classification interval maximization and the classification error minimization. The Lagrange method of multipliers will be applied to change the above quadratic planning problem into its dual form:

$$\begin{aligned} \max \quad & W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \\ & \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned}$$

At this point, the linear classifier derived is:

$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^l \alpha_i y_i (x \cdot x_i) + b \right)$$

where α_i ($i = 1, 2, \dots, l$) is the Lagrange multiplier, and each α_i corresponds with one training sample. In fact, there are many α_i 's strictly equaling to 0, and only those samples corresponding to the non-0 α_i will affect the results. Such samples are called support vectors. It can be seen that the training results of support vectors are only related to support vectors and are not related to non-support vectors.

2.2 Nonlinear SVM

For a nonlinear classification problem, the samples will be mapped onto higher-dimension feature spaces through the nonlinear mapping function $\varphi(x)$, thus transforming the problem into a linear classification problem within the higher-dimension feature spaces. The kernel function $K(x, y) = \varphi(x) \cdot \varphi(y)$ will be introduced to replace the inner product operation of the two vectors, and therefore the computation complexity is not increased. At this point, the problem can be transformed into the following quadratic planning problem [1]:

$$\begin{aligned} \max \quad & W(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, l \\ & \sum_{i=1}^l \alpha_i y_i = 0 \end{aligned}$$

At this point, the linear classifier derived is:

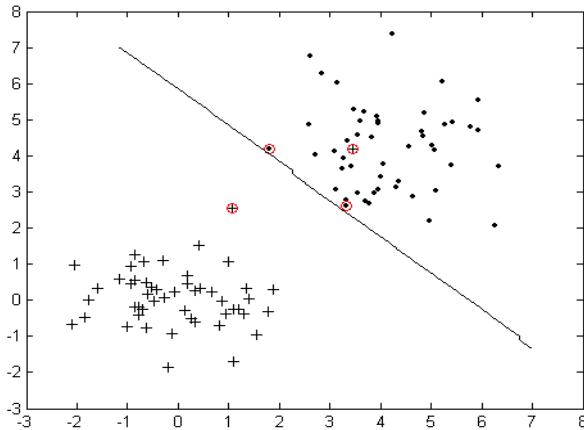
$$f(x) = \operatorname{sgn} \left(\sum_{i=1}^l \alpha_i y_i K(x, x_i) + b \right)$$

3 The Iteration Algorithm of Pre-extracting Support Vectors

3.1 The Geometrical Explanation of the SVM

Fig. 1 is a graphic representation of a SVM in a two-dimensional space. In Fig. 1, the circled dots represent support vectors, and the solid line running across the figure represents the dividing line, thus the distribution of the support vectors can be seen clearly. Viewed directly from the geometrical perspective, the support vectors are composed of the converging area dots and outliers of the two categories of sample dots, and the outliers seriously affect the performance of the classifier. Therefore, if most of the support vectors can be pre-extracted and the outliers can be eliminated in the meantime before training, then the training speed and generalization ability of the SVM will be increased.

First, the linear SVM will be taken as an example to pre-extract support vectors, and then the entire process will be generalized to the nonlinear SVM.

**Fig. 1.** Two-dimensional classification SVM

3.2 Pre-extracting Support Vectors of the Linear SVM

(1) The definition used in this study

Definition 1: The definition of the class center is as follows:

The mean feature of the i th sample ($i \in \{1, 2\}$) is the class center of this class of samples, that is:

$$m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_j^{(i)}$$

(2) The rationale of the class center-based algorithm of pre-extracting support vectors

In Fig. 2, two straight lines will be made perpendicular to the line connecting the two class centers, then most of the support vectors will fall within the band area, and the outliers can be effectively eliminated in the meantime. Based on such rationale, the iteration algorithm will be adopted and iteration will be conducted successively so as to attain the purposes of pre-extracting support vectors and eliminating the outliers.

Theorem 1: The function of the hyperplane which runs through the class center m_i of the i th sample ($i \in \{1, 2\}$) and is perpendicular to $\overrightarrow{m_1 m_2}$ is as follows:

$$\Pi_i \quad \overrightarrow{m_1 m_2} \cdot (x - m_i) = 0, \quad i = 1, 2$$

Namely:

$$\Pi_i \quad \frac{1}{n_2} \sum_{k=1}^{n_2} x \cdot x_k^{(2)} - \frac{1}{n_2 n_i} \sum_{k=1}^{n_2} \sum_{l=1}^{n_i} x_k^{(2)} \cdot x_l^{(i)} - \frac{1}{n_1} \sum_{k=1}^{n_1} x \cdot x_k^{(1)} + \frac{1}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} x_k^{(1)} \cdot x_l^{(2)} = 0, \quad i = 1, 2$$

Theorem 2: The distance between the two hyperplanes of Π_1 and Π_2

$$d(\Pi_1, \Pi_2) = \left\| \overrightarrow{m_1 m_2} \right\| = \sqrt{\frac{1}{n_2^2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} x_k^{(2)} \cdot x_l^{(2)} - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} x_k^{(1)} \cdot x_l^{(2)} + \frac{1}{n_1^2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} x_k^{(1)} \cdot x_l^{(1)}}$$

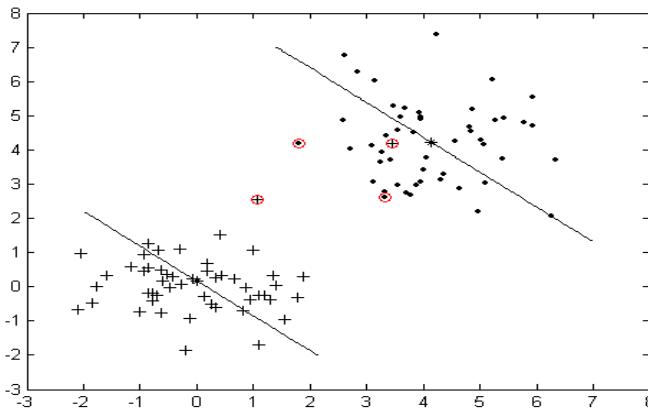


Fig. 2. Graphic representation of the two-dimensional class center-based algorithm of pre-extracting support vectors

Theorem 3: The distance from any sample dot x to the hyperplane Π_i

$$d(x, \Pi_i) = \frac{\left| \frac{1}{n_2} \sum_{k=1}^{n_2} x \cdot x_k^{(2)} - \frac{1}{n_2 n_i} \sum_{k=1}^{n_2} \sum_{l=1}^{n_i} x_k^{(2)} \cdot x_l^{(i)} - \frac{1}{n_1} \sum_{k=1}^{n_1} x \cdot x_k^{(1)} + \frac{1}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} x_k^{(1)} \cdot x_l^{(2)} \right|}{\sqrt{\frac{1}{n_2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} x_k^{(2)} \cdot x_l^{(2)} - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} x_k^{(1)} \cdot x_l^{(2)} + \frac{1}{n_1^2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} x_k^{(1)} \cdot x_l^{(1)}}}$$

Theorem 4: For any sample dot x , if $d(x, \Pi_1) + d(x, \Pi_2) \leq \|m_1 m_2\|$, then x will be within the band area formed between the two hyperplanes of Π_1 and Π_2 ; otherwise, x will be out of the area.

(3) The steps of the class center-based algorithm of pre-extracting support vectors

Suppose the upper limit of the quantity of the preset pre-extracting vectors as n_0 and the sample dot sets of the two classes of samples as S_1 and S_2 .

Step 1: If $|S_1| + |S_2| \leq n_0$, then stop, otherwise, go on to the next step;

Step 2: Get the class centers m_1 and m_2 of S_1 and S_2 , and then through the class centers of the two classes, make two hyperplanes perpendicular to the line connecting the two class centers;

Step 3: Get the dot sets S_1 and S_2 of the two classes of samples within the band area; turn to Step 1.

3.3 Pre-extracting Support Vectors of the Nonlinear SVM

For the nonlinear classification problem, the samples will be mapped onto higher-dimension feature spaces through the nonlinear mapping function $\varphi(x)$, thus transforming the problem into a linear classification problem within the higher-dimension

feature spaces. Similar to the definition in the linear problem, we define the class center of the i th class of samples ($i \in \{1, 2\}$) as $m_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varphi(x_j^{(i)})$. Because at this point it is impossible to know the specific form of the function $\varphi(x)$, we can introduce the kernel function $K(x, y) = \varphi(x) \cdot \varphi(y)$ to substitute for the inner product operation of the two vectors, similar to the linear classification problem, then the following conclusion can be made:

Theorem 5: In higher-dimension feature spaces, the function of the hyperplane running through the class center m_i of the i th class of samples ($i \in \{1, 2\}$) and perpendicular to $\overrightarrow{m_1 m_2}$ is as follows:

$$\Pi_i : \overrightarrow{m_1 m_2} \cdot (\varphi(x) - m_i) = 0, \quad i = 1, 2$$

Namely:

$$\Pi_i : \frac{1}{n_2} \sum_{k=1}^{n_2} K(x \cdot x_k^{(2)}) - \frac{1}{n_2 n_i} \sum_{k=1}^{n_2} \sum_{l=1}^{n_i} K(x_k^{(2)} \cdot x_l^{(i)}) - \frac{1}{n_1} \sum_{k=1}^{n_1} K(x \cdot x_k^{(1)}) + \frac{1}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} K(x_k^{(1)} \cdot x_l^{(2)}) = 0,$$

$$i = 1, 2$$

Theorem 6: In higher-dimension feature spaces, the distance between the two hyperplanes Π_1 and Π_2 is as follows:

$$d(\Pi_1, \Pi_2) = \left\| \overrightarrow{m_1 m_2} \right\| = \sqrt{\frac{1}{n_2^2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} K(x_k^{(2)} \cdot x_l^{(2)}) - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} K(x_k^{(1)} \cdot x_l^{(2)}) + \frac{1}{n_1^2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} K(x_k^{(1)} \cdot x_l^{(1)})}$$

Theorem 7: In higher-dimension feature spaces, the distance from any sample dot x to the hyperplane Π_i is as follows:

$$d(x, \Pi_i) = \sqrt{\left| \frac{1}{n_2} \sum_{k=1}^{n_2} K(x \cdot x_k^{(2)}) - \frac{1}{n_2 n_i} \sum_{k=1}^{n_2} \sum_{l=1}^{n_i} K(x_k^{(2)} \cdot x_l^{(i)}) - \frac{1}{n_1} \sum_{k=1}^{n_1} K(x \cdot x_k^{(1)}) + \frac{1}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} K(x_k^{(1)} \cdot x_l^{(2)}) \right|^2 - \left(\frac{1}{n_2^2} \sum_{k=1}^{n_2} \sum_{l=1}^{n_2} K(x_k^{(2)} \cdot x_l^{(2)}) - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_2} K(x_k^{(1)} \cdot x_l^{(2)}) + \frac{1}{n_1^2} \sum_{k=1}^{n_1} \sum_{l=1}^{n_1} K(x_k^{(1)} \cdot x_l^{(1)}) \right)^2}$$

Theorem 8: In higher-dimension feature spaces, for any sample dot x , if $d(x, \Pi_1) + d(x, \Pi_2) \leq \left\| \overrightarrow{m_1 m_2} \right\|$, then x will be within the band area formed between the two hyperplanes of Π_1 and Π_2 ; otherwise, x will be out of the area.

Similar to the algorithm of pre-extracting support vectors in the linear classification problem, in higher-dimension feature spaces, the pre-extracted vectors can be used to replace the entire training sample set, thus increasing the training speed of the SVM.

4 Simulation Experiments

Example 1: Instance of the linear SVM

Two classes of two-dimensional normal samples $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ were randomly generated. The distribution of the first class of samples was $N(0, 0, 1, 1, 0)$,

and the distribution of the second class of samples was $N(5,5,1,1,0)$. In the course of training, the linear kernel function was adopted, the compromise parameter was taken as $C=10$, and the upper limit of the quantity of the vectors to be extracted was set as 20. After that, 10 times of training were conducted and the mean value was adopted. Fig. 3 is a graph randomly picked during the course of experimenting. In Fig. 3, “+” and “•” represent the two classes of samples respectively; the circled dots represent the support vectors obtained through the standard SVM algorithm; the dots with “*” superimposed represent the pre-extracted support vectors. From Fig. 3 it can be seen that the pre-extracted vectors contained all the support vectors, and meanwhile, most of the non-support vectors were eliminated, thus reducing the training samples and extremely shortening the training time. Table 1 presents the training results.

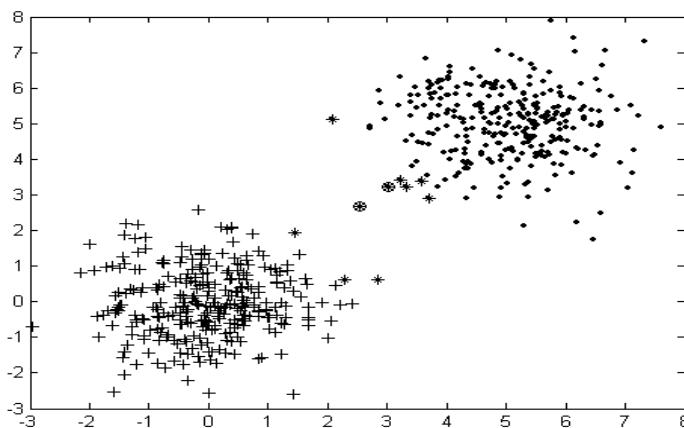


Fig. 2. The pre-extraction of support vectors with linear division

Table 1. Comparison of the classifying performance of the linear SVM

Training algorithm	Training samples (cases)	Parameter	Pre-extracted vectors (cases)	Support vectors (cases)	Training time (s)	Test samples (cases)	Recognition rate (%)
Standard SVM	600	-	-	3	866	1400	99.93
Pre-extracting + SVMs	600	-	11	3	0.015	1400	99.93
Reference [9]	600	$\mu=0, D=0$	13	3	0.015	1400	99.93

Example 2: Instance of the nonlinear SVM

Two classes of concentric circle samples $x=R \cos \theta, y=R \sin \theta$ were randomly generated, in which $\theta \in U[0, 2\pi]$. The radius of the first class of samples is $R_1 \in U[0, 5]$, and the radius of the second class of samples is $R_2 \in U[5.1, 10]$. In the course of the training, the radial basis kernel function was adopted. The upper limit of the quantity of the vectors to be extracted was set as 150. After that, 10 times of

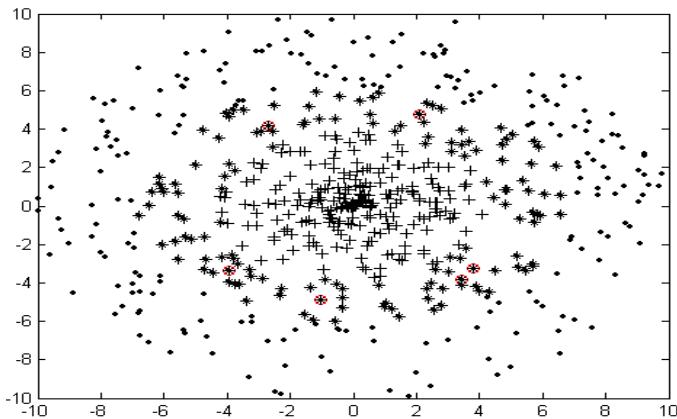


Fig. 3. The pre-extraction of support vectors with nonlinear division

Table 2. Comparison of the classifying performance of the nonlinear SVM

Kernel function and Parameter	Training algorithm	Training samples (cases)	Pre-extracted vectors (cases)	Support vectors (cases)	Training time (s)	Test samples (cases)	Recognition rate (%)	
Radial basis kernel function	Standard SVM	600	—	19	1086	1400	99.86	
$\sigma=5$	Pre-extracting + SVMs	600	—	123	6.359	1400	99.71	
C=100	Reference [9]	600	$\mu=0.15$, $D=10$	86	20	1.289	1400	99.64

training were conducted and the mean value was adopted. Fig. 4 is a graph randomly picked during the course of experimenting. From Fig. 4 it can be seen that the pre-extracted vectors contained all the support vectors, and meanwhile, most of the non-support vectors were eliminated. Table 2 presents the training results.

Example 3: Testing with UCI data

The three data sets of Breast cancer Wisconsin, Ionosphere, and Waveform were selected from the UCI database. The Breast cancer Wisconsin data set was composed of 30 feature attributes and 1 class attribute, and it could be seen as a dichotomous problem. In the course of training, the linear kernel function was adopted, with the compromise parameter set as $C = 20$, the upper limit of the quantity of the vectors to be pre-extracted set as 50. The Ionosphere data set was composed of 34 feature attributes and 1 class attribute, and it could be seen as a dichotomous problem. In the course of training, the radial basis kernel function was adopted, with the parameter set as $\sigma = 4$, the compromise parameter set as $C = 10$, the upper limit of the quantity of the vectors to be pre-extracted set as 50. The Waveform data set was composed of 21 feature attributes and 1 class attribute, and it could be seen as a three-class problem. Therefore, its 0-1 class, 0-2 class, and 1-2 class were selected respectively to construct SVMs for training. In the course of training, the radial basis kernel function

was adopted, with the parameter set as $\sigma = 10$, the compromise parameter set as $C = 10$, and the upper limit of the quantity of the vectors to be pre-extracted set as 250. After that, 10 times of training were conducted and the mean value was taken. The training results are presented in Table 3.

Table 3. Testing the proposed algorithm of pre-extracting support vectors with the UCI data sets

Data sets	Training algorithm	Training samples (cases)	Parameter (cases)	Pre-extracted vectors (cases)	Support vectors (cases)	Training speed (s)	Test samples (cases)	Recognition rate (%)
Breast cancer	Standard SVM	200	-	-	18	13.203	369	94.85
Wisconsin	Pre-extracting + SVMs	200	-	46	9	0.250	369	92.95
in	Algorithm in reference [9]	200	$\mu=0.1, D=10$	118	18	3.094	369	94.85
Ionosphere	Standard SVM	200	-	-	67	5.266	151	97.35
	Pre-extracting + SVMs	200	-	42	30	0.125	151	96.69
	Algorithm in reference [9]	200	$\mu=0.1, D=10$	185	67	4.11	151	97.35
Waveform	Standard SVM	1000	-	-	201	1598	2304	91.58
0-1 Class	Pre-extracting + SVMs	1000	-	247	151	13.938	2304	90.45
	Algorithm in reference [9]	1000	$\mu=0.02, D=5$	507	193	126.109	2304	91.06
Waveform	Standard SVM	1000	-	-	189	1349	2353	92.01
0-2 Class	Pre-extracting + SVMs	1000	-	238	153	10.828	2353	91.84
	Algorithm in reference [9]	1000	$\mu=0.02, D=5$	469	185	94.406	2353	92.01
Waveform	Standard SVM	1000	-	-	133	1203	2343	95.35
1-2 Class	Pre-extracting + SVMs	1000	-	202	126	6.704	2343	95.26
	Algorithm in reference [9]	1000	$\mu=0.02, D=5$	302	133	20.813	2343	95.31

5 Conclusion

Through the study and analysis of the mechanism and geometrical characteristics of the SVM, it was found that the support vectors were composed of the converging area

dots and the outliers of the two classes of sample dots, therefore the class center-based iteration algorithm of pre-extracting support vectors was proposed. By means of selecting artificial data and the three data sets from the UCI database, the proposed algorithm was analyzed and verified, and it was found that this algorithm could substantially reduce the quantity of training samples and enhance the speed of computation. Besides, the following conclusions were drawn: (1) When the support vectors in the training samples only accounted for some of the training samples, namely, when there was a considerable redundancy in the training samples, the proposed algorithm was effective, and the larger the redundancy, the more apparent the effect; (2) Compared with the algorithm in reference [9], in the course of pre-extracting support vectors, the algorithm proposed in this study only requires setting the upper limit of the quantity of the support vectors to pre-extracted, without requiring setting other parameters, thus it is convenient for operation and the precision obtained thereby is comparable; (3) When the two classes of samples can be classified completely linearly or completely nonlinearly, in terms of the quantity of the vectors pre-extracted and the classifying performance, the algorithm proposed in this study and that in reference [9] is comparable; in addition, when there is intersection between the two classes of samples, namely, when there is noise in the class attributes, the quantity pre-extracted in the algorithm proposed in this study is less than that in reference [9] and the training speed of the proposed algorithm is faster, but in terms of classifying performance, the two algorithms are comparable.

References

1. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer (1995)
2. Osuna, E., Freund, R.: An Improved Training Algorithm for Support Vector Machines. In: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, pp. 276–285 (1997)
3. Platt, J.C.: Sequential Minimal Optimization: A fast algorithm for training support vector machines, pp. 169–182. The MIT Press, Cambridge (1998)
4. Cao, L.J., Keerthi, S.S., et al.: Parallel Sequential Minimal Optimization for the Training of Support Vector Machines. *IEEE Transaction on Neural Networks* 17(4), 1039–1049 (2006)
5. Li, H.-L., Wang, C.-H., Yuan, B.-Z.: An Improved SVM: NN-SVM. *Chinese Journal of Computers* 26(8), 1015–1020 (2003)
6. Ye, F., Luo, J.-Q., Yu, Z.-F.: An Improvement SVM Learning Algorithm with Parallel Processing. *Microelectronics & Computer* 26(2), 40–43 (2009)
7. Zheng, C.H., Jiao, L.C.: Fuzzy Pre-extracting Method for Support Vector Machine. In: Proceedings of the 1st International Conference on Machine Learning and Cybernetics, pp. 2026–2030 (2002)
8. Jiao, L.-C., Zhang, L., Zhou, W.-D.: Pre-extracting Support Vectors for Support Vector Machine. *Acta Electronic Sinic* 29(3), 383–386 (2001)
9. Li, Q., Jiao, L.-C., Zhou, W.-D.: Pre-extracting Support Vector for Support Vector Machine Based on Vector Projection. *Chinese Journal of Computers* 28(2), 145–152 (2005)
10. Wu, Q., Liu, S.-Y., Du, Z.: Fuzzy Support Vector Machine Method Based on Border Vector Extraction. *Pattern Recognition and Artificial Intelligence* 21(3), 332–337 (2008)

Research on Inbound Tourist Market of Liaoning Province Based on Tourism Background Trend Line

Wei Guo, Yixue Liu, and Hong Guo

Tourism Department School of Economics & Management, Yanshan University,
Qinhuangdao, China
guowei@ysu.edu.cn

Abstract. To achieve the purposes of optimizing the inbound tourist market in Liaoning province, the article analyzed the development trends of inbound tourist market in the future, which based on the statistics data of inbound tourists of Liaoning province from 1996 to 2009. It used tourism background trend line model to establish the tourism background trend line equation of tourist arrivals and inbound tourism income in Liaoning province. Following it drew tourism background trend chart. Then, the article used the natural extension of tourism background trend line to predict the future. The results show that the prediction error is small. The inbound tourist market of Liaoning province will maintain steady growth in the next five years. Finally, it proposed strategies to develop inbound tourist market in Liaoning province.

Keywords: Inbound Tourist Market, Tourism Background Trend Line, Trend Forecast, Liaoning Province.

1 Introduction

The theory and empirical studies made by Archer (1996), Sinclair (2003), Araujo Lnn (2009) have concluded that the tourism can accelerate economic development, the contribution of tourists' spending will increase in double (Murphy, 1986).Tourism is an association industry, it can boost the development of the relevant industries and optimize the economic structure. Due to the special geographical position of Liaoning province, the inbound tourist market plays an important role in the overall tourism market. Through reading the research literatures of tourism market, the literatures mostly analyze inbound tourism flow with the date in a certain time. This article used tourism background trend line model to predict the future development trend of the inbound tourist market in Liaoning province, with the aim to optimize the tourism industry.

The principle and the conceptual model of tourism background trend line or BTL were proposed by Sun Gennian. Tourism background trend line refers to the inherent tendency equation without serious impact and interference (Sun Gennian,1998). Background trend line is based on time as independent variable. Its equation can transform the development trend of the past into a natural extension of the future, so it has the prediction function. In addition, it can evaluate special events in the tourist destination.

2 Tourism Background Trend Line Model

Tourism background trend line equations are different from the previous simple linear simulation(Witt Stephen R,1983). They are generalized in several forms, including four basic form: Linear growth, exponential line growth, logical line growth, the sinusoidal line growth and other compound forms, such as straight - index line, straight - logic line, Index - line, Index - logic line, logic - sine line and so on(Sun Gennian,1998). To establish the optimal background trend line equation, this study mainly drew the time series statistical figure and selected optimal equation.

3 The Establishment of the Background Trend Line towards Inbound Tourist Market in Liaoning Province

3.1 The Selection and Processing of Data

In order to be able to understand the development trend about inbound tourist market in Liaoning province more accurately in the future, this study collected the date of tourist arrivals and inbound tourism income. Then, it used tourism background trend line to analyze and predict the development trend.

This study established the tourism background trend line model according to the statistic data from 1996 to 2009 collected in China statistical yearbook. In the beginning of the article, tourism background trend line is defined under no serious impact. The data of tourist arrivals and inbound tourism income in special year need to be revised before the establishment of tourism background trend line in Liaoning province. For example, the date in 2003 should be revised for the impact of SARS. This research used linear interpolation correction to revise the statistics in the special year. The method shows as follows: Firstly, drawing the time series statistical figure based on the statistics. Secondly, searching for the linear interpolation of the starting point (n_a) and terminate (n_b) by observing the chart. Thirdly, revising the statistic data of the special year because of the important events according to the interpolation equation: $Y_n = Y_a + (n - n_a) \times d$. Y_a represents the starting point, Y_b represents the terminate value. The calculation formula of d : $d = (Y_b - Y_a) / (n_b - n_a)$. The revised data of tourist arrivals and inbound tourism income showed in Table 1.

Table 1. The revised data of tourist arrivals and inbound tourism income in Liaoning province

Indicator	Endpoint of linear interpolation		Tolerance value	Interpolation equation	The revised data 2003
	Starting point	Terminate point			
Tourist arrivals	2002	2004	75 690.5	$Y_{2002} + (n - n_{2002})d$	1 005 119
Tourism income	2002	2004	3 130	$Y_{2002} + (n - n_{2002})d$	58 151

3.2 The Establishment of the Background Trend Line Equation

After drawing the time series statistical figure with the revised statistic data, the second step is to select a suitable trend line equation according to several background trend line equations. The third step is to determine the parameter and the eventual background trend line equation through data fitting with Marlab7.0. In the trend line equation listed below, t stands for the time variable from 1996 to 2009, which followed by taking $t = 1, 2, 3$,

The background trend line model of tourist arrivals in Liaoning province

The dynamic trend line of tourist arrivals in Liaoning province was in sine wave on the basis of exponential growth. The background trend line equation:

$$y = 237696.34 \exp 0.178t + 13987.58 \sin(9.90t + 20.26) \quad (1)$$

The correlation coefficient was 0.975.

The background trend line model of inbound tourism income in Liaoning province

The background trend line of inbound tourism income in Liaoning province was in sine wave on the basis of a linear growth. The equation:

$$y = 98088.58t - 402310.73 + 1001502.84 \sin(0.0095t + 2.72) \quad (2)$$

The correlation coefficient was 0.979.

3.3 Tourism Background Trend Chart

After the establishment of the background trend line equation, this research drew tourism background trend figure in order to react the relationship between background trend line and the original statistical line about the tourist arrivals, inbound tourism income in Liaoning province, also including the aim to verify the accuracy of background trend line model.

The background trend chart of tourist arrivals in Liaoning province

Tourist arrivals are very important indicators to react the level of tourism development. This study made a combined graph based on the statistic date of tourist arrivals about 14 years in Liaoning province. Refer to Fig. 1. The smooth curve in the chart is the background trend line. The natural extension part of the statistical line is the forecast in the next few years.

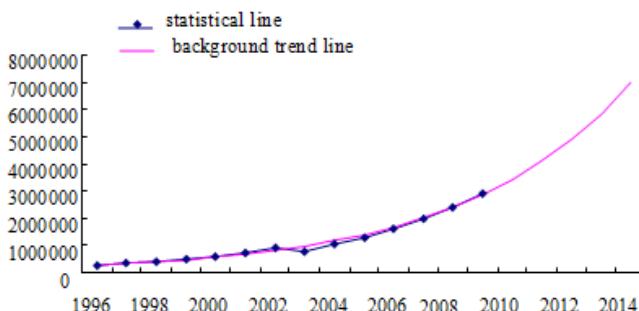


Fig. 1. The background trend chart of tourist arrivals in Liaoning province

From Fig. 1, it can be seen that the statistical line lied below the background trend line and formed a "hollow valley" in and around 2003 because of SASR. Due to unexpected events in special year, the statistical line doesn't fit with the background trend line, which is a normal phenomenon. Tourist arrivals in Liaoning province will maintain a steady growth trend in the next five years.

The background trend chart of inbound tourism income in Liaoning province.

The income of inbound tourism is another important indicator to response the development of tourism. This article drew a figure. Refer to Fig. 2.

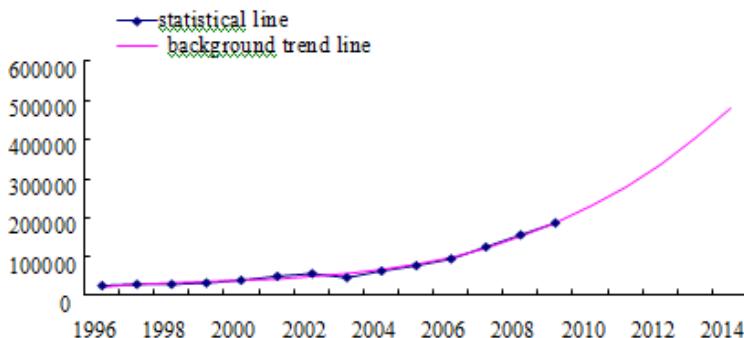


Fig. 2. Background trend chart of inbound tourism income in Liaoning province

From Figure 2, it also can be seen that the statistical line lied below the background trend line and formed a "hollow valley" because of the impact of SASR. The inbound tourism income will increase in the next few years.

4 The Forecast of Inbound Tourist Market in Liaoning Province

In this part, it used tourism background trend equation to forecast tourist arrivals, inbound tourism income (Refer to Table 2).

Table 2. The predicted value of tourist arrivals and inbound tourism income in Liaoning province

Year Indicator	2010	2011	2012	2013	2014
Tourist arrivals (person-time)	3 414 523	4 098 154	4 889 533	5 833 710	6 989 235
Inbound tourism income (million U.S. dollars)	3 414 523	4 098 154	4 889 533	5 833 710	6 989 235

Compared the predicted value with the actual data about tourist arrivals and inbound tourism income in 2010, it got that the prediction error was from -1% to 3%. The forecast which based on tourism background trend line was accurate. It can develop the inbound market clearly with the predicted value.

5 The Improvement and Optimization of Inbound Tourist Market

5.1 The Establishment of Tourism Brand Image in Liaoning Province

Combination of the tourism resources in Liaoning province, firstly, Liaoning province should stimulate the establishment of three major tourist areas, including Liaodong peninsula, the central and western of Liaoning, Secondly, establish representative tourism product system in Liaoning province. Then, it can construct tourism brand on behalf of the whole image of Liaoning province.

5.2 Improving Inbound Tourist Market through Marketing Campaign

Liaoning province can propagate its tourism brand image and special tourism products through integrating tourism resources, regional cooperation and joint promotion. It can also organize festival activities to attract tourists, propagate the tourism slogan and logo through web sites, media, the press, etc. Meanwhile, it can build information platform to expand inbound tourist market.

5.3 Optimizing the Tourism Industry System and the Structure of Inbound Tourism Consumption

Liaoning province should improve each aspect of tourism, including food, housing, transportation, travel, shopping, and entertainment and so on. Then, optimizing the structure of the tourism product and promoting international tourism consumption structure in Liaoning province. The aim is to promote the development of the inbound tourist market in harmony and stability.

6 Conclusion

Based on the statistics data from 1996 to 2009 and the tourism background trend line model, this study established the tourism background trend line equation and tourism background trend chart of Liaoning province. Then, it used tourism background trend equation to forecast tourist arrivals, inbound tourism income. The results show that the prediction error is small. The tourist arrivals, inbound tourism income would maintain steady growth in the next five years. Finally, this article proposes strategies to develop inbound tourist market in Liaoning province.

Acknowledgement. The author would like to acknowledge the great support of the national social science fund project-“Study on mechanism, innovative mode and conformity strategy of tourism promoting economic growth” (Grant Numbers:09BJY087) and the study topics of social science development in Hebei Province in 2010(Regional economic association fund projects, Topic Numbers: 201005004).

References

1. Lnn, A.: Partnership and Regional Tourism in Brazil. *Annals of Tourism Research* 29, 1138–1164 (2009)
2. Sun, G., Zhou, R., Ma, L.: Impact of five Important Events on China Provincial Inbound Tourism in 2008: An Analysis in High-Resolution Based on the TBL Model. *Scientia Geographica Sinica* 31, 1437–1446 (2011)
3. Shu, J., Sun, G.: TBT- Line and Crisis Evaluation of 9 Intercontinental Tourist- Source Markets of US. *Statistics & Information Forum* 25, 41–46 (2010)
4. Li, Z.: An Research on the Foundation and Application of the Background Trend-line of Domestic Tourism in China. *Statistics & Information Forum* 24, 62–65 (2009)
5. Ma, Q., Sun, G.: On the Growth of Tourism Industry based on TBCM- A Case Study of Inbound Tourism in Guilin. *Economic Geograph* 29, 1402–1408 (2009)
6. Sun, G.: Foundation and Significance of Background Trend Line of Tourists from Abroad in China. *Scientia Geographica Sinica* 18, 442–448 (1998)
7. Li, N., Sun, G.: Foundation and Significance of Background-trend-line of Xi'an City's Abroad Tourism. *Journal of Northwest University* 28, 339–344 (1998)
8. Sun, G.: Foundation of Tourism Background-trend-line of 6 major Tourism-Source-markets of China. *Systems Engineering-theory & Practice* 20, 140–143 (2000)
9. Sun, G.: The dynamic tendency and fluctuation cycle of inbound tourism in Xián city. *Journal of Northwest University* 31, 514–517 (2001)
10. Witt Stephen, R.: Modeling and forecasting demand in tourism. *Recreation Research Review* 6, 25–29 (1983)

Research on Influence of Tourism on Economy

Dan Luo and Wuzhong Zhou

Institute of Tourism and Landscape Architecture, School of Arts, Southeast University,
Nanjing 210096, China
dana_luo@126.com

Abstract. Tourism can enhance economic growth and can also benefit from the growth. In this paper, in order to identify the mode of tourism-economy relationship, a multi-dimensional time series of both tourism and economy was analyzed with principal component analysis (PCA). Our results revealed that, in 1989, most of the European countries shared a similar mode, in which tourism had a great contribution to both GDP and employment, and the capital investment was at a high level. In 2000, the developing countries advanced so much that their tourism-economy mode was similar to that of the developed countries. In 2011, the modes diversified. The United States showed a distinct mode. The mode of China gradually resembles that of the United States. In China, the government and individual tourism spending is increasing.

Keywords: tourism-economy mode, principal component analysis, clustering, dynamics.

1 Introduction

There is an intimate and complex relationship between tourism and economy. In 2011, more than 9% of global gross domestic product (GDP) was contributed by the tourism industry [1]. The tourism can promote the economy. However, in some countries, such as in Korea, the economy drives the growth of the tourism [2-3]. The tourism is also sensitive for the changes in economic development. In this economic crisis, the international tourism maybe be the latest sectors that feel the effects of the global recession [1,3].

Due to the differences in tourism resource, economic environment and government policy, the causal relationship between tourism and economy differs from one country to another [3-4]. A study showed the stock price of tourism agents is long-run equilibrium with GDP in China; but causality tests between hotels, airlines, and travel agents and GDP showed a mixed causality result, indicating a complex interaction between tourism and economy [5]. The causality is bidirectional in some countries [6]. Moreover, for one country, the tourism's impact on economy varies with time. The tourism industry is in crying need of information and knowledge for decision making and for strategies to effectively respond to the current situation. Thus, investigating the interplay is extremely important for governments in establishing strategies for both tourism and economy.

The tourism-economy relationship can be represented with the statistic quantitative features related to both tourism and economy [1,3-4]. The features include the tourism's contribution to GDP and employment, visitor export, capital investment, and so on. They vary with time. How to integrate the multi-dimensional time series data and capture the dynamics of the association between tourism and economy is a great challenge in resolving the mode of tourism development [7-8]. Panel unit root tests is widely used in investigating the causal relationship between the tourism and the economy [3,9]. However, it needs tourism real receipts per capita, which is difficult to count. Moreover, the result of the unit root tests is hard to visualize, thus limits the further analysis.

In this paper, a principal component analysis (PCA)-based method was employed in identifying the dynamics of the tourism's impact on economy in eighteen countries from 1989 to 2011. The results showed that the tourism industry positively associated with the economic growth; the mode of relationship between the two is dynamic. The mode of China gradually resembles that of the United States.

2 Dataset and Method

The dataset related to the tourism-economy relationship was retrieved from the website of World Travel & Tourism Council (WTTC) with "Economic Data Search Tool" [1]. It is a 20-dimension data from 1989 to 2011 for 18 countries, representing with $F(T_i, C_j, P_k)$, where T , C and P indicate time, country and feature, respectively; $i=1$ to 24, $j=1$ to 18, $k=1$ to 20. Here, the features include: tourism direct contribution to GDP and its growth (P_1 and P_2); tourism total contribution to GDP and its growth (P_3 and P_4); tourism direct contribution to employment and its growth (P_5 and P_6); visitor exports and its growth (P_7 and P_8); domestic tourism spending and its growth (P_9 and P_{10}); government and individual tourism spending and its growth (P_{11} and P_{12}); internal tourism consumption and its growth (P_{13} and P_{14}); leisure tourism spending and its growth (P_{15} and P_{16}); business tourism spending and its growth (P_{17} and P_{18}); capital investment and its growth(P_{19} and P_{20}).

At each time point, PCA was performed and the two fore principal components ($PC1$ and $PC2$) were extracted for each country to quantitatively represent the tourism-economy mode. The details of the PCA procedure were as followed [7-8,10]. For time t , given matrix A ($A=F(t, C_j, P_k)$), the first step is to standardize A, resulting Z (eq.1),

$$Z_{jk} = \left(a_{jk} - \bar{a}_k \right) / s_k, \quad j=1,2 \dots 18, \quad k=1,2 \dots 20 \quad (1)$$

Where \bar{a}_k indicates mean of the k -th feature in 18 countries, and s_k is standard deviation. Then calculate covariance matrix R (eq.2).

$$R = ZZ^T \quad (2)$$

At last, resolve the eigenvectors and the eigenvalues of R , and extract the two fore principal components ($PC1_t$ and $PC2_t$).

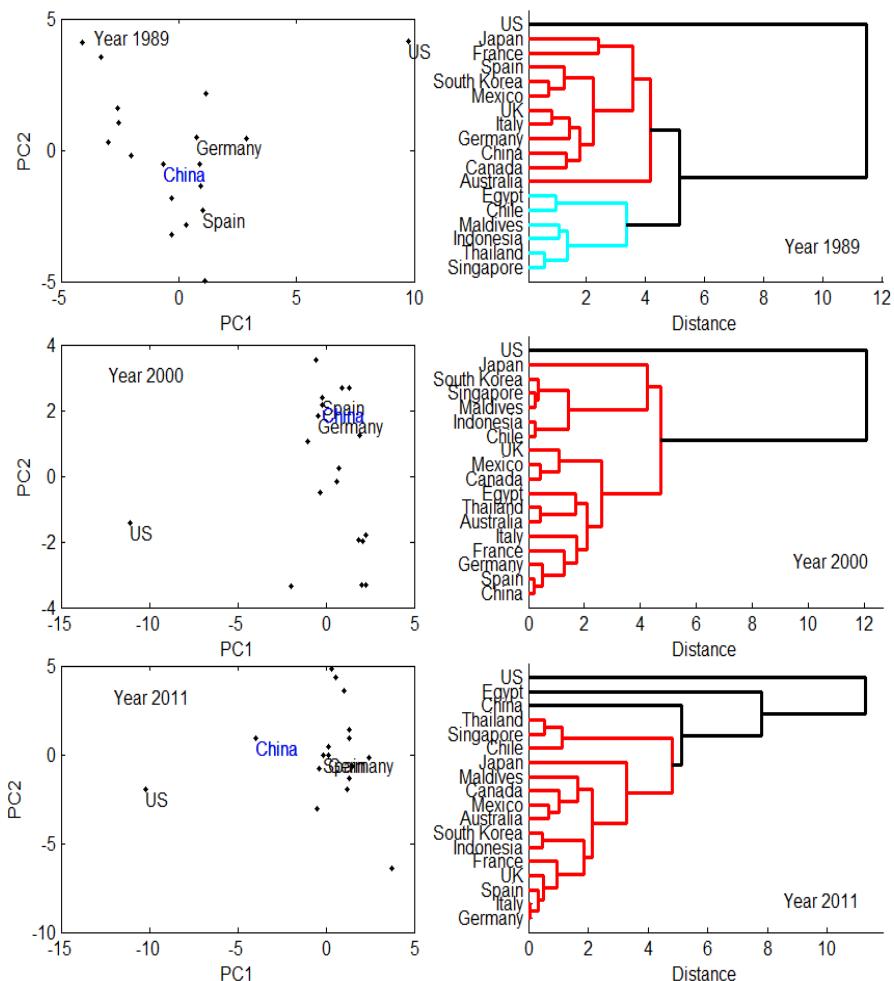


Fig. 1. The tourism-economy mode differs among the countries; the left subplots, the scores of the eighteen countries on the two fore principal components (PC1 and PC2) in 1989, 2000 and 2011, respectively; the right subplots, the hierarchical clustering using PC1 and PC2. PC1 and PC2 have an averaged contribution ratio of more than 76.52%

In order to investigate the dynamics of the tourism-economy mode, the scores of each country on the two components are examined at each time (year). The similarity of the tourism-economy mode of any two countries is assessed with the Euclidean distance of the two countries at the plane of $PC1_t$ and $PC2_t$. Moreover, hierarchical clustering is used to classify the eighteen countries with the $PC1_t$ and $PC2_t$.

3 Results and Discussion

Despite the challenges from terrorist attacks and health scares to natural disasters, tourism still shows a growth over the past decade. The progress of tourism promotes

the economy and the development of the economy also show an enhancing for the tourism, showing an interplay mode. The mode differs from one country to another. We compared the modes through clustering (Fig.1). In $PC1$ and $PC2$ plane, the United States (USA) shows a great difference from other countries from 1989 to 2011. The mode of China has a trend of closing to USA.

In 1989 (Fig. 1), most of the European countries, including Spain, Germany, France, the United Kingdom (UK) and Italy, clustered together, indicating the countries share a similar tourism-economy mode. From Fig. 2 (left), it can be deduced that the countries has a high level in features P_3 , P_7 , P_9 , P_{13} , P_{15} , P_{17} and P_{19} , but has no significant growth, showing an stable development mode. We notice the capital investment (P_{19}) is high in countries of the cluster. Probably, this is one of factors that cause the lack of growth in both tourism and economy. A recent literature demonstrated that capital investment has side effect that can reduce the true benefits for host communities in the case of small tourism economies [11]. This suggests that policy makers should carefully assess the impacts of capital investment in a broader socio-political framework in attracting foreign investment into tourism development.

Another characteristic of the clustering map in 1989 is that tourism resource-rich countries, including Egypt, Maldives and Thailand, are in a same cluster. The countries in the cluster are located in Asia and Africa. Their economies are developing. The clustering for 1989 shows that there is a separation between the developing countries and the developed countries.

In 2000, the clustering shows that the developing countries and the developed countries cluster together, suggesting that the some of the developing countries advance so much that their tourism-economy mode became more similar with that of the developed countries.

In 2011 (Fig.1), the tourism-economy modes diversify. The obvious clustering is not observed. The modes of China, US, and Egypt differ from others. China is featured by the increasing of government and individual tourism spending (P_{11} and P_{12}) (Fig.2 right). This is consistent with the WTTC's report that stated that China was the second biggest spender in both US and Europe markets [1, 3].

In order to analyze the mode of tourism-economy in china, USA and Spain are chosen to compare. USA is one of advanced economies. The USA dominates global tourism spending, accounting for more than a quarter (26%) of the business travel market in 2011 [1]. Spain is the fourth largest tourism economy country [4]. Fig.3 shows the dynamics of the distances of China-USA and China-Spain in plane of $PC1$ and $PC2$ from 1989 to 2011. A significant trend is China becomes more closely to USA, especially after 2000. Meanwhile, the distance between China and Spain increases from 1989 to 2011, suggesting a growing difference in tourism-economy modes between the two countries. The results indicate that in China tourism significantly contributes to economy, but it is not the only reason accounting for the growth of the economy. Tourism enhances economy, but it also benefits from the economic growth.

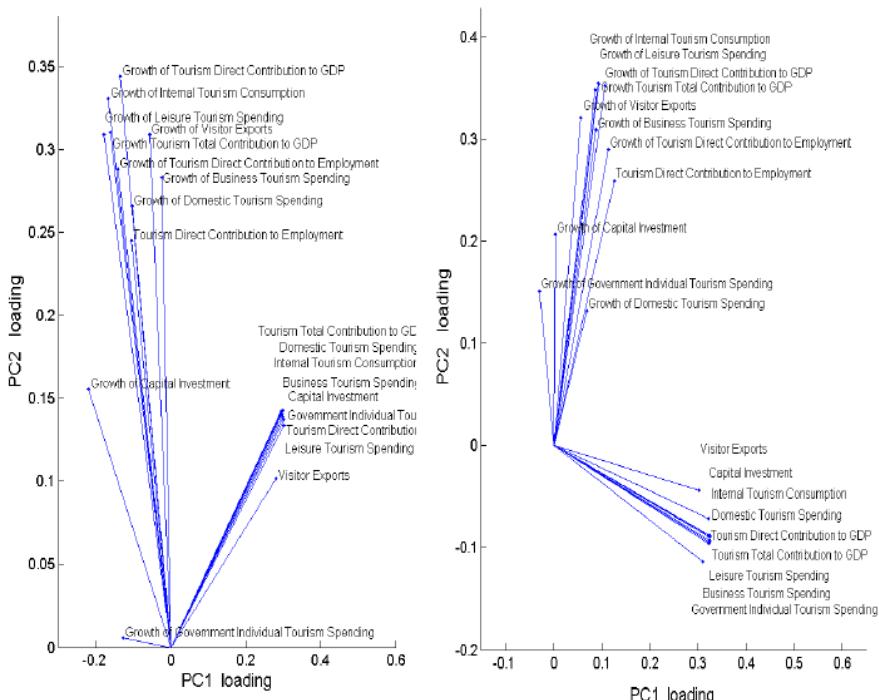


Fig. 2. The distribution of the features in the fore two principal components (PC_1 and PC_2) plane, the left subplot is for 1989, and the right one is for 2011

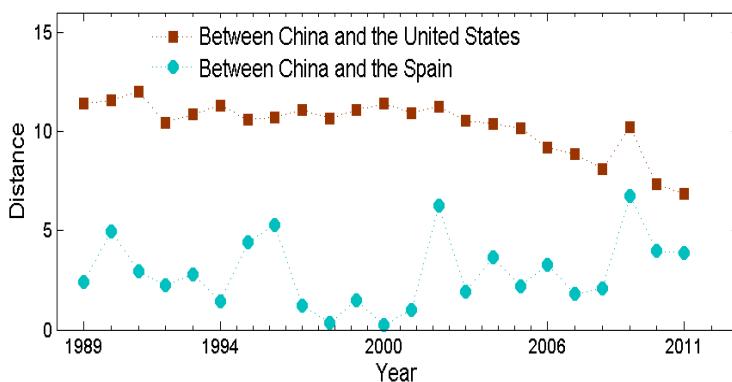


Fig. 3. Dynamics of the distances of China-the United States and China-Spain in plane of PC_1 and PC_2 from 1989 to 2011

In Fig.3, another important finding is that curve of China-US distance is smoother than curve of China-Spain distance, which can also be quantitatively represented with the coefficients of variation (CV) of the two curves, namely 13.06% and 61.19% for China-US distance and China-Spain distance, respectively. The results indicate that

China's development of both economy and tourism are more tightly associated with that of US. At the time points of the Asian financial crisis in 1997, extraordinarily serious flood in 1998 and SARS epidemics in 2003, the curve of China-US distance does not show significant oscillation, but the curve for China-Spain distance varies greatly at the time points. This suggests that the affect of the events on tourism and economics are synchronously and closely associated in China and US.

4 Conclusion

We analyzed the dynamics of the tourism-economy mode in eighteen countries. In the past two decades, the modes are changed and diversified. The United States showed a distinct mode comparing with other countries, and the mode of China became more similar to that of the United States. In China, the government individual tourism spending is increasing. We also demonstrated a technique of processing the multi-dimension and multi-sample time series data.

References

1. World Travel & Tourism Council (WTTC), <http://www.wttc.org/>
2. Figini, P., Vici, L.: Tourism and growth in a cross section of countries. *Tourism Economics* 16(4), 789–805 (2010)
3. Tang, C.H., Jang, S.: The tourism–economy causality in the United States: A sub-industry level examination. *Tourism Management* 30, 553–558 (2009)
4. Seetanah, B.: Assessing the dynamic economic impact of tourism for island economies. *Annals of Tourism Research* 38(1), 291–308 (2011)
5. Chen, M.H.: Interactions between business conditions and financial performance of tourism firms: evidence from China and Taiwan. *Tourism Management* 28, 188–203 (2007)
6. Evangelia, K.: Economic Impact of Tourism on Greece's Economy: Cointegration and Causality Analysis. *Internat. Res. J. Finance & Economics* 79, 79–85 (2011)
7. Kadous, M.W.: Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. PhD thesis, University of New South Wales (2002)
8. Jolliffe, I.T.: Principal Component Analysis, ch. 1. Springer, New York (2004)
9. Lee, C.C., Chang, C.P.: Tourism development and economic growth: a closer look at panels. *Tourism Management* 29, 180–192 (2008)
10. Singhal, A., Seborg, D.E.: Clustering multivariate time-series data. *J. Chemometrics* 19(8), 427–438 (2005)
11. Li, S., Tsui, Y.: Foreign Investment in Tourism: The Case of Macao as a Small Tourism Economy. *Tourism Geographies* 12(2), 173–191 (2010)

Hydrodynamic Characteristics Experiment of Vegetation Flexible Dam

Feng Liu¹, Liang Ma², Liang-Liang Chen², and Guang-Hui Wei²

¹ Division of Development and Programming, Xinjiang Agricultural University,
Urumqi City, Xinjiang Province, China

² School of Water Resources and Civil Engineering, Xinjiang Agricultural University,
Urumqi City, Xinjiang Province, China
wlmqlf1999@163.com

Abstract. By the indoor flume model experiment, the paper analyzed systematically the influence of three different base slopes of vegetation "flexible dam" to the water flow under the cross layout. The experiment showed that the base slopes is more big after planting trees, the discharge is more loudly unknown to the reflection on backwater. Pass to descent the analysis that plant trees to the different discharge in same base slopes at the same time in front and back mean velocity in section, explained vegetation dam can let up dam front goodly with the average current velocity of the inside of dam, and turn the current velocity peak value even. The vegetation dam can also nicely cut down bed to shear in response to the dint and let up starting of sediment to drag dint, the sediment doesn't easily start but deposition in the gutters. And the sea buckthorn has a very strong divide grows, since make the grove covered up by the sediment to press, will also born new root and side, continue upward growth, undertake to block a sand task. Finally from analyzed to plant trees theoretically in front and back the variety of the sediment deposition.

Keywords: Vegetation "Flexible Dam", Base Slope, Sediment.

1 Introduction

In 1995, professor Bi Ci-fen, Li Gui-fen and Yu Zhuo-de pointed out that: Soft rock channel runoff capacity of soil erosion on coarse, torrential rain and flood erosion characteristics of a single main proposed the use of sea buckthorn plant "flexible dam" to intercept gully grit [1]. After several years, which have been part of the coarse solid in the gully, reducing the grit into the Yellow River, but also to the ecological environment in these areas to improve. The so-called sea buckthorn plant "flexible dam" (hereinafter referred to as plant dam) is the main canals in the channel or tank, at a certain spacing and line spacing, planting a number of sea buckthorn plants. After investigation, soft rock are also common in Xinjiang, and its living environment even worse, resulting in relatively more coarse, so the more urgent need to address.

Plants dam for the western region of the, especially in the sand for the Xinjiang region, combating desertification and soil erosion control work provides a new idea.

Xinjiang Agricultural University, Institute of Water Resources and Hydraulic Engineering Laboratory in 2002 and 2003 studies of indoor plant dam water tank model tests, which examine the plants in the ranks of the aligned and staggered dam layout, different planting density on the canals the impact of water[2,3]. To further study the mechanism of plant dam-fixing, the group carried out the indoor water tank model test, the water flows through the plant dam on the flow pattern when a detailed observation.

2 Experiment Overview

2.1 Test Equipment

This test is carried out within the glass tank. tank length 22m, width 0.4m, high 0.5m, with the first tank inlet valve and static pool, the pool is located hydrostatic energy dissipation grid plate to stabilize the flow. Tank has adjustable tailgate end, after falling into the water tank via the tail gate. To intercept sediment to the stream, sediment in the water tank placed a sieve. The Tank is equipped with active stylus holder, read the water level can be measured, the bed elevation and so on. Tank top with the scale of the track, for moving the stylus holder and read the measured cross-section location. Would sink the scale where the zero cross-section of the track record for the SC0+00 section, 1m cross-section where the scale denoted SC1+00, according to the location of the remaining cross-section scale and so on. The bottom end of the tank consists of a fixed hinge, the first, located under the bottom end of an active hinge, for regulating the tank bottom slope.

The water level measured by needle which has accuracy of 0.1mm, flow measured by using LGY-III-type multi-function smart flow meter.

2.2 Choice of Model Tree

In order to simulate natural sea buckthorn tree truthfully on the impact of water and sediment, taking into account the resistance of trees and deformation similar to the experiment with decorative green plastic tree as a model tree. About 10cm high plastic around the whole tree, trunk diameter of about 2mm, the upper part of branches and leaves, top of the tree maximum horizontal width of 7cm, and has a certain flexibility. Model tree roots can be used for fixed circular hole, test the model tree in a socket fixed on the grid. Model tree lined water out of the water-blocking effect of preferred best $6\text{cm} \times 9\text{cm} \times 9\text{cm}$, planting density of $248/\text{m}^2$.

2.3 Model Scale

This study aimed at Xinjiang's rivers and more coarse problem. Through field investigation, the adult height of sea buckthorn is about 1m to 2m; test plastic tree full height of about 10cm or so, consider the convenience to study, mainly based on the model to determine the height of the tree than the scale length, 1:10 normal model

selection, using the median particle size Non-uniform model of natural sand as bed load sediment.

3 Test and Analysis of Experiment Phenomenon

The first test when water flows through the plant dam the flow pattern was observed in detail. Observations show that: the sink for a certain arrangement of the plant behind the dam, through its whole form a dam, and interfering with the original water flow patterns within the tank, causing the dam tour stop up the water level is high, the maximum depth of plant dam tour starting at the dam end of SC8+00 cross-section at. Plants within the water depth along the dam long gradually reduced, and with wavy, each row of plants, there are clear of the vortex area. Dam downstream from the dam end of the plant at 5cm depth down to the minimum depth of the tank to form a contraction section, then the downstream water depth is increasing, and gradually returned to the water depth before planting. As the plant itself is flexible, so the water flows through the plant dam, the plant bending deformation, the flow direction is along the lines of flow and deformation in its equilibrium position after a slight swing, and looks like a comb over the same neat . This is due to water trees is different from the solid-liquid two-phase flow often say the one hand, the flow is from the stems, and through the trees, the trees on them by the block; the other hand, trees in the water power, the occurrence of the corresponding deformation and the equilibrium position to produce the reciprocating vibration, the vibration can be part of the turbulent flow from the energy[4]. Wang Cun[5] with a large soft leaves of plants in the sink to do the test, also observed bending and oscillation of vegetation and that water flows through the flexible vegetation, in essence, is a moving boundary problem.

From the viewpoint of fluid mechanics, every tree within the plant dam is actually a very complex boundary conditions flow around the cylinder, the resistance increases from the tree-shaped water-blocking area and material resistance, in their common generated under a row of trees blocking effect on the water, and thus the formation of the whole plant dam effect, making the tank within the plant section of the dam tour kinetic energy decreases, resulting in changes in water flow, resulting in plant dam backwater tour, behind the dam erosion the phenomenon.

In this experiment, the average water depth is the water tank at the depth of the vertical axis. Flow rate of 9.08L/s under the bottom slope is different before and after the cross-section of average water depth of planting trees along the distribution curve. From the figures 1 and figures 2, we can see clearly in the sink 8m scale SC8+00 dam at the upstream section, before planting, after planting depth is much larger than the water depth, resulting in a backwater, and the flow of water inside the tank after planting in the SC8+00 to SC9+00 sharp decline in cross-section between the water depth, water depth in the section behind the dam SC9 +00 bottoming out. The depth and gradually began to recover.

The first three trees under the average depth of the bottom slope changes along the way or less, just as the tank bottom slope becomes steeper, the water depth decreases, but the magnitude of reduction is not large. After planting the size of the average depth of the bottom slope along the way influenced changes in the upstream dam SC8+00 section, three at the end of the lower slope of the reduction was larger

backwater; in SC8+00 downstream section, the average water depth than close to. After planting, 1/150 and 1/100 when the bottom slope to the curve is similar to the average depth is only 1/150 to 1/100 when the water depth should be large, indicating that 1/150 when the bottom slope upstream backwater plant dam caused high, and 1/50 when the average depth of the bottom slope to the curve of the bottom slope and the first two are different.

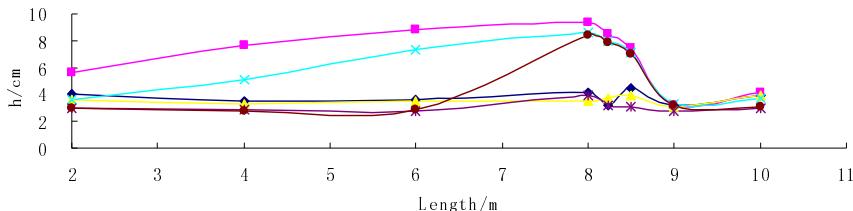


Fig. 1. Distribution diagram of cross-section average water before and after planting at the same discharge and the different base slope

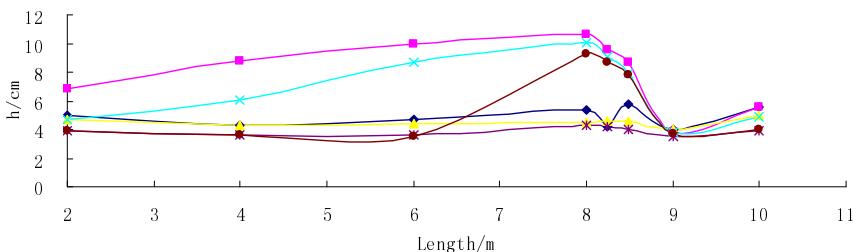


Fig. 2. Distribution diagram of cross-section average water before and after planting at the same base slope and the different discharge

Experiments found that: the same flow rate, 1/100 when the bottom slope is not submerged trees, 1/50 when the bottom slope was all drowned, and in 1/50 when the bottom slope upstream of the backwater section of the water there were many more large waves, the greater the flow, the higher peaks. This is because as the bottom slope becomes steeper, upstream flow rheology of the slow jet, and the plant due to plant dam effects, larger water depth, water flow is slow, so slow transition here from the jet stream the formation of wavy water jump, increasing the flow turbulence, which will affect the role of plant dam. So we choose plant "flexible dam" site should be considered more moderate bottom slope of the lot, so as to play a good silt results.

As can be seen, when the same flow rate, 1/150 after the end of the dam slope planting depth amplitude, followed by the 1/100 at the end of the slope, and finally 1/50 bottom slope. Section in SC8+00 dam downstream of the plant behind the dam cross-section of three kinds of bottom slope and water depth when the amplitude approaches, the cross section close to the SC10+00 shows the depth of amplitude with tree planting around the bottom slope decreases.

It can be seen from the Fig.1, in a different flow, the cross-section average velocity along the trend is the same. After planting the various flow chart scale that SC8+49 at

8.49m upstream average velocity cross-section than before planting a small, but also reduces the rate was relatively large. Section in SC9+00 after planting at a flow rate of greater than or close to the velocity before planting. It can better explain the decrease of plant dam and the average flow velocity, and peak flow homogenization. Therefore, the natural channel of the plant can reduce the sediment dam starting frequency, and allows the sediment has been deposited starting to play sand, sediment effect.

From the above plant "flexible dam" flow characteristics of the analysis shows that water plants have a good blocking effect. It can reduce the flow velocity dam, dam water depth increases. Shows a certain spacing and line spacing perpendicular to the flow direction staggered planting sea buckthorn, the role of energy dissipation can be achieved. The principle is: When the gully in the planting sea buckthorn plant, the water to the sea buckthorn trees pop when shot by plants, branches, leaves and block the diversion, the concentration-stream dispersion, reduction of shear flow on the gully stress, reduce sediment delivery ratio, thus blocking the flow of large amounts of sediment in the mix, so that part of the sediment deposited in the dam tour. Also, because the sea buckthorn has a strong bushy, even trees buried by sediment pressure, will produce new adventitious roots and lateral branches continue to grow up, assume the task of sediment [6-10]. This also demonstrates that plants grown after the main tank, because the blocking effect of the water, the dam significantly reduced the flow rate will result in the upstream part of the larger sediment particles in the sediment deposition, bed rise, while plants with near both sides of the river's edge and increase the flow velocity on the slope, slope erosion increased, resulting in the wide-shallow river-based development. If sediment-laden flow, the dam at the plant will have sediment deposition, the conditions are conducive to promote high plant growth. Conversely, the high growth of plants, turn to intercept more of sediment, makes it into a benign cycle track.

4 Conclusion and Discussion

The existence of plant "flexible dam" interference with the original flow pattern, the tank water before the flow field and tree planting has been significantly different. Each plant trees within the dam, is a very complicated boundary conditions the flow around the problem, in their joint action produces a row of trees blocking effect on the water, and thus the formation of the whole plant dam effect, making the plants inside the tank dam tour section kinetic energy to overcome plant resistance generated by the dam on the flow, resulting in plant dam backwater tour, behind the dam erosion phenomenon. Backwater upstream of the dam in the plant area after planting the cross-section of the flow velocity is less than before planting, Therefore, plants in the natural channel upstream of dam sediment can reduce the starting frequency, and make the start of the sediment has been deposited, so as to achieve sand, sediment effect.

Roughness and average boundary shear stress in all significant changes before and after planting. After planting trees in front of channel roughness than the large, while the average boundary shear stress smaller than the pre-planting. Roughness after

planting to reduce the flow rate increases rapidly, the roughness of the decline can be regarded as plant stems between its deformation and over-current, the resistance decreased due to the shape. Flexible, especially submerged plants over-current, after the plant at the top of lodging, easy to form a smooth water area, increases the over-current capability. However, as long as there are trees, and its roughness on the river with a significant increase effect. It reduces the drag force starting sediment, so sediment is not easy to start and deposited in the ravine.

Changes in the plant dam sediment deposition patterns, and deposition of sediment particle sorting carried out. After planting a single sediment than the average velocity is significantly smaller before planting, after planting, mainly with the smaller bed shear stress related. Different bottom slope, different time flow and sediment transport under different sediment deposition patterns are relatively similar, relatively smooth surface of the body deposition, but the thickness of sediment deposition, location and vertical lengths are different. The existence of plant dam not only for the same cross-section of the table bottom sediment were sorting, and deposition of the sediment vertical sorting of particles also, showing changes in backwater areas of the reservoir sedimentation law.

Plant dam-fixing mechanism of the theoretical analysis. When the gully in the planting sea buckthorn plant, the water to the sea buckthorn trees pop when shot by plants, branches, leaves and block the diversion, the concentration-stream dispersion, reduction of water flow on the channel bed shear stress, reducing the sediment delivery ratio, thus blocking the flow of large amounts of sediment in the mix, so that part of the sediment deposited in the dam tour. Also, because the sea buckthorn has a strong bushy, even trees buried by sediment pressure, will produce new adventitious roots and lateral branches continue to grow up, take sediment tasks.

Acknowledgments. This research work was supported by the Xinjiang Agricultural University Pre-school Funding Project(XJAU201008). We give the most heartfelt thanks to the reviewers for their great contributions on the early manuscript.

References

1. Bi, C.-F., Yu, Z.-D., Li, G.-F.: Control soil erosion in soft rock areas of new ideas. In: Second Cross-Strait exchanges Water Science and Technology Symposium, Taipei (1996)
2. Qiu, X.-Y., Abuduwaili, Cheng, Y., et al.: Study on effect of plant "flexible dam" to the flow. Water Resources and Hydropower Technology 34(9), 62–65 (2003)
3. Hu, C.-H., Hui, Y.-J.: Nullah sediment-laden flow movement mechanics and statistical laws. Science Press, Beijing (1995)
4. Shi, B., Cao, S.-Y.: Plant Sand Dynamics. Qingdao Ocean University Press, Qingdao (2000)
5. Wang, C.: Vegetated river flow studied (Master thesis). Hohai University, Nanjing (2003)
6. Liu, F., Yan, J., et al.: Plant "flexible dam" experimental study on the mechanism of sediment. River Sediment (6), 34–37 (2006)
7. Xiao, J., Zheng, Z., Wang, J.: Analysis on suspended and bed-load ratio of river sediment for South Xinjiang. Water Resources and Hydropower Engineering (4), 30–34 (2009)

8. Sun, D.-P., Li, B., Tong, T.: Analysis on Environmental Effect of River Sediment. *Journal of Irrigation and Drainage* (6), 11–15 (2010)
9. Wu, X.-G., Liu, G.-S., Cheng, W.-L.: 3D numerical simulation of sediment in the Hangzhou Bay based on FVCOM. *Hydro-Science and Engineering* (6), 11–15 (2011)
10. Cui, P., Ge, Y.-G., Lin, Y.-M.: Soil Erosion and Sediment Control Effects in the Three Gorges Reservoir Region. *China Journal of Resources and Ecology* (4), 37–40 (2011)

Solar Energy Product Based on Variable Length Gene Expression and Intuition Models

Ting-ting Zhao¹, Xiao-peng Wei², Zhi-qiu Xia², Hong-jiang Wang¹, and Jun-tie Xia²

¹ Department of Information Engineering, Shenyang Institute of Engineering,
Shenyang, China

² Key Laboratories of Advanced Design and Intelligent Computing,
Dalian University, Ministry of Education, Dalian, China
zhaottingting2006@163.com

Abstract. A new creative design approach is put forward based on variable length gene expression and intuition models. The number of product requirements, the gene length expressing requirements, the structure of the product, and the correlation matrix are varied with individuation of customer requirements of the product. By research on the calculation mechanisms of dynamic variety, the approaches of variable length gene expression are proposed. The intuition process reflects that the thinking of human brain produces qualitative change. An approach of intuition simulation is used by Hebb learning law, Hopfield neural networks and crossover and mutation. The calculating models and the calculating formulas for the creative design are put forward. Finally, An example for the creative design of a solar energy portable lamp is given. The better results are obtained in the creative design.

Keywords: Gene Expression, Intuition Models, Creative Design.

1 Introduction

There are some works on the product gene, the study focused on the creative design stage. Gábor Renner [1] gave a summary to this problem. Tingting Zhao[2] extended single-object to multi-object optimization models and established a multi-object calculation model from function requirements to structures by the correlation matrixes, then proposed a uniform optimal calculation model for conceptual design and solved it out by genetic algorithm. As time goes on, new requirements are to be added in the set, old requirements are removed from the set and new requirements bring about new structures. In the mean time the length of gene chains for expressing product functions is changed which means that the content of initial product information cannot be used in the database, so the product must be renewed design. A lot of time and human resources are wasted.

Creative design process includes cognitive process of mankind. The famous professor Qian Xue-sen has divided the noetic science into three parts, which are logical thinking, thinking in images and inspiration thinking[3]. The theories of the

traditional artificial intelligence based on logic have made a great progress[4]. People also study thinking in images deeply[5-10]. The third thinking (intuition, insight and inspiration) is an advanced cognitive form. R. Cooper [4] put forward the extreme importance of intuition, insight and inspiration, and psychological phenomenon of inspiration thinking have been imitated by means of observing the difference between the experts and new recruit. Professor Qian Xue-sen[3] pointed out the breakthrough point of researching noetic science is thinking in images (inspiration thinking). Zhao Ting-ting [5] expounded in detail the importance of intuition in creative thinking. But it is difficult to deal with and made use of intuition by symbols because of the characteristic of itself. The studies of intuition are just the beginning.

This paper proposes a new design approach of variable length gene and intuition generation. Consequently, the experiment results for a solar energy portable lamp are completed.

2 Variable Length Gene Expression

2.1 Gene Expression of Product

It is very important for conceptual design of complicated products by the quantitative expression of design process, which can make automatic achievement of conceptual design easy. So each element of products such as function, behavior, structure and carrier can be named as design elements in the design process. Gene expression of design elements can be defined as following:

Definition 1 Gene expression of design element: Give a design element such as function, behavior or structure as an ordered set F_k , $k=1,2,\dots,n$, and let it a n -dimensional vector. According to its position, establish a n -dimensional vector f , $f=(f_1,f_2,\dots,f_n)$, whose value equals to 0 or 1, that is:

$$f_k = \begin{cases} 1, & \text{included in } F_k \\ 0, & \text{excluded in } F_k \end{cases}, k = 1, 2, \dots, n \quad (1)$$

Then the vector f is named as gene expression of design element for the given product, which includes requirements gene expression, function gene expression, structure gene expression, behavior gene expression and carrier gene expression etc.

2.2 Variable Length Gene Expression

The changes of ordered set F_k , $k=1,2,\dots,n$, can result in the changes of the gene expression of each vector. However the changed gene expression can be figured out by the previous gene expression. So if we want to add or delete some element from a certain position, shown as Fig.1 and 2, original gene position should be moved horizontally.

(1) Add the design elements of s number after the position i

Suppose the initial gene expression vector of a design element is $(b_1, b_2, \dots, b_i, \dots, b_n)$, and the final gene expression vector is $(c_1, c_2, \dots, c_i, \dots, c_{i+s}, \dots, c_{n+s})$, which is shown in Fig.1, so that:

$$c_j = \begin{cases} b_j & \text{when } j \leq i \\ \text{adding elements} & \text{when } i < j \leq i+s \\ b_{j-s} & \text{when } j > i+s \end{cases} \quad (2)$$

(2) Delete the design elements of s ($s < n$) number after the position i

Suppose the initial gene expression vector of a design element is $(b_1, b_2, \dots, b_i, \dots, b_n)$, and the final gene expression vector is $(b_1, b_2, \dots, b_i, \dots, b_{n-s})$, which is shown in Fig.2, so that:

$$d_j = \begin{cases} b_j & \text{when } j \leq i \\ b_{j+s} & \text{when } j > i \end{cases} \quad (3)$$

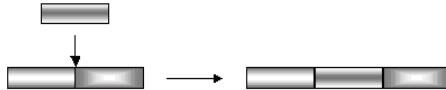


Fig. 1. Change of position when gene segment is added



Fig. 2. Change of position when gene segment is deleted

3 The Calculating Models of Intuition

The following gives the calculating models of intuition. Simulate the process of experience accumulation with the Hebb's law. Enlightenment condition is marked with vector. Associative process is achieved with Hopfield neural networks. Mutation process chooses one sort of the three operators, which are crossover operators, mutation operators and crossover mutation operators. Show in figure1.

Associative process and mutation process are achieved in parallel in the cognitive models, but they are achieved in the calculating models because of the limit of parallel calculating ability of computers.

The calculating process is the following:

(1) Implementing experience by Hebb's law

Suppose p is vectors: $\{u^1, u^2, \dots, u^p\}$, $u^i \in \{-1, +1\}^n$. Then we can get weight matrix:

$$W = \frac{1}{n} \sum_{k=1}^p u^k u^{kT} \quad (4)$$

The vectors u^1, u^2, \dots, u^p can be as fixed points of the network under some certain conditions. When a vector is input, the networks can find a stable state having the shortest Hamming distance from this vector, and stop iteration.

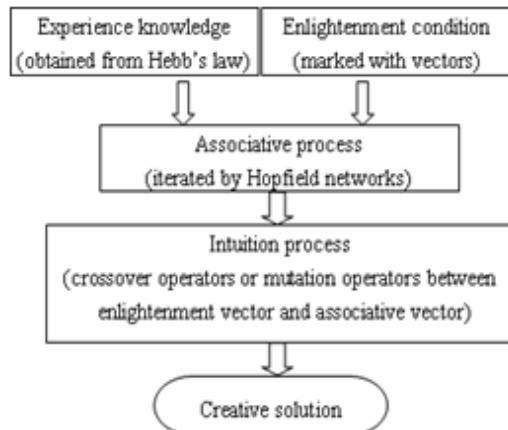


Fig. 3. The calculation models of intuition

(2) Achieving associative memory by the Hopfield networks

The synaptic matrix W is acquired from Hebb's law. There are n neurons and θ_i is its threshold. v_i is the state of neuron i , whose value must belong to the set $\{-1, 1\}$. The network adopts parallel working mode.

Consider the enlightenment vector u^0 as the original vector of Hopfield networks, and have the networks iterated. The following is its formula:

$$\begin{cases} v_i^{(0)} = u_i^0 \\ v_i^{(k+1)} = \text{sgn} \left[\sum_{j=1}^n w_{ij} v_j^{(k)} - \theta_i \right] \end{cases} \quad (5)$$

The state of network will not change any longer after some limiting moment, that is $v_i^{(k+1)} = v_i^{(k)}$, $i = 1, \dots, n$, then the network is in the stable state. If this state corresponds to one u^l ($l = 1, \dots, p$) of the p deposited prototypes in the network, that is $v_i^{(k)} = u^l$, then we consider the vector u^l as the associative vector obtained by the vector u^0 .

(3) Crossover or mutation

The crossover process between the vector u^0 and the vector u^l show the merge between the enlightenment vector and the associative vector, and mutation of them is used to show the mutation of thinking. Then the resulted vector can produce intuition.

a. Achieving coding methods

According to different questions, we select different coding method, which includes binary coding, floating-point coding, integer coding etc.

b. Achieving crossover

According to different coding methods, we select different crossover operator. For example, use one-point crossover or multiple point crossovers about binary coding. Use arithmetic crossover about floating-point coding etc.

c. Achieving mutation

We also choose different mutation operator according to different states. About binary coding we change 0-1 values of one or several genes according to the mutation rate. About floating-point coding we use non-uniform mutation operator.

4 The Calculating Process

The calculation process is the following:

(1) Gene expression of product

Gene expression of design elements can be defined as formula 1.

(2) Variable length gene expression

Add the design elements of s number after the position i as formula 2.

Delete the design elements of s ($s < n$) number after the position i as formula 3.

(3) Hebb's law process

Hebb's law process as formula 4.

(4) The associative memory

The associative memory by the Hopfield networks as formula 5.

(5) Crossover or mutation operators

We select different crossover operator according to different coding methods. For example, use one-point crossover or multiple point crossovers about binary coding. Use arithmetic crossover about floating-point coding etc.

We also choose different mutation operator according to different states. About binary coding we change 0-1 values of one or several genes according to the mutation rate. About floating-point coding we use non-uniform mutation operator.

5 Application Case

In order to realize the creative design in such a way above, a solar energy portable lamp is utilized as the experiment instance of product model. The main components of the instance of the solar energy portable lamp include: PV panel, battery, controller, load, tilt angle, shape of box, and location of parts. PV panel, battery, controller, and load are the static part of the product gene. Show in Figure 4. The process of design need not consider their change. Tilt angle, shape of box, and location of parts is the dynamic part. The process of design need consider these.

The shapes of box have split style box, one-piece box. The product gene expression is (1, 0), (0, 1).

$(1, 0), (0, 1)$ are taken as sample of Hebb's law. We can get weight matrix:

$$W = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

We choose that the convenience is as the enlightenment vectors. The input vectors of the networks are $(0, 1)$. By associative memory of the Hopfield networks we can get that $(0, 1)$ is a stable attractor of networks. At last the one-piece box is chose. Show in Figure 5.

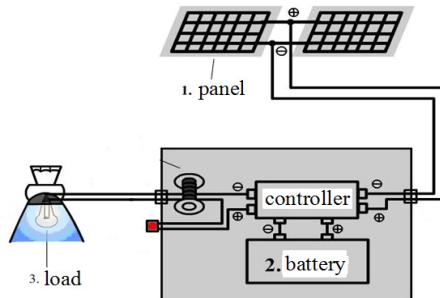


Fig. 4. The main components of the solar energy lamp

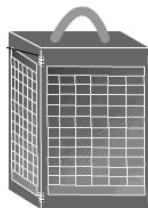


Fig. 5. One-piece box

We used variable length gene expression approach for the enlightenment vectors. The lighting area is added to the enlightenment vectors. By crossover and mutation operators between the associative vectors and the enlightenment vectors, so we used folding design approach. Show in Figure 6.

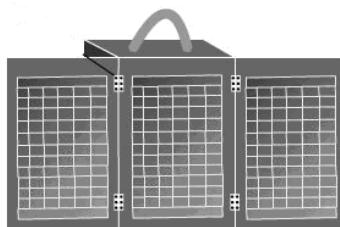


Fig. 6. Folding box

Because tilt angle need equal to optimal tilt angle, We take optimal tilt angle as the enlightenment vectors. By crossover and mutation operators between the associative vectors and the enlightenment vectors, we can arrive at the result which is optimal tilt angle folding design approach. Show in Figure 7.

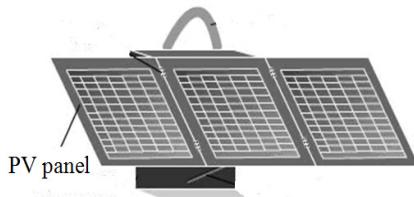


Fig. 7. Optimal tilt angle design

6 Conclusion

According to creative design based on variable length gene expression and intuition models, this paper gets the conclusions as follows:

- (1) A new variable length gene expression model is presented and its sequential gene manipulation is proposed based on the new variable length gene.
- (2) The mutual relationships are simulated among experience, intuition, association and enlightenment by Hebb learning law, Hopfield neural networks and crossover and mutation manipulation.
- (3) A new creative design approach is put forward based on variable length gene and intuition models. An instance of a solar energy portable lamp is given to evaluate the performance of the proposed variable length gene and intuition models.

Acknowledgments. The research is partially supported by National Natural Science Foundation of China (No.50975033, No.60875046), and Education Office of Liaoning Province (No. LT2010074).

References

1. Renner, G., Ekárt, A.: Genetic algorithms in computer aided design. Computer-Aided Design 35, 709–726 (2003)
2. Zhao, T., Wei, X.: Gene expression-based multi-objective optimization Conceptual Design of Product Structure. Chinese Journal of Mechanical Engineering 41, 102–107 (2005)
3. Zhao, G., Wang, J., Lu, M.: Noetic sciences studies. Chinese People University Press, Beijing (1999)
4. Cooper, R., Fox, J., Farrington, J., Shallice, T.: A systematic methodology for cognitive modeling. Artificial Intelligence 85, 3–44 (1996)
5. Zhao, T., Wei, X., Zheng, H.: Intuitive simulation method based on associative memory. World Congress on Intelligent Control and Automation 3, 2030–2032 (2010)
6. Wei, X., Zhao, T.: Approach to creative design based on intuitive model. Computer Integrated Manufacturing Systems 11, 7–11 (2010)

7. Olcer, A.I.: A hybrid approach for multi-objective combinatorial optimization problems in ship design and shipping. *Computers and Operations Research* 32, 86–97 (2011)
8. Yi, P., Wei, X., Zhao, T.: Creative design on surface of products combining NURBS with genetic algorithm. *Computer Integrated Manufacturing Systems* 14, 221–225 (2008)
9. Torstenfelt, B., Klarbring, A.: Conceptual optimal design of modular car product families using simultaneous size, shape and topology optimization. *Finite Elements in Analysis and Design* 43, 1050–1061 (2011)
10. Hsiao, S.-W., Tsai, H.-C.: Applying a hybrid approach based on fuzzy neural network and genetic algorithm to product form design. *International Journal of Industrial Ergonomics* 35, 411–428 (2010)

Study on the Logistics Efficiency of Three Northeast Provinces Based on Three-Stage DEA

Sun Xu^{1,2}

¹ Key Laboratory of Bionic Engineering of Ministry of Education Jilin University,
Changchun Jilin 130025 China

² Jilin Technology college of Electronic Information Jilin, 132021 China
sunxusk@sohu.com

Abstract. This paper established an index system of input and output based on the survey of the logistics present situation in three northeastern provinces, according to the index of input and output ,dealing with the data of the logistics input and output of 18 regions in three northeastern provinces ,estimated the logistics efficiency of three northeast provinces ,using three-stage DEA, the results indicates that:(1)the pure technical efficiency and the scale efficiency of the three northeastern provinces are all lower than 0.9, and there is a big rise in them;(2)from the pure technical efficiency, the results of three northeastern provinces show the annual increasing trend on the whole, and there are no obvious differences in regional logistics efficiency;(3)from the scale efficiency, the scale efficiency of three northeastern provinces show the annual increasing trend on the whole, but there is a big gap between the scale efficiency level and the ideal efficiency level. It indicates that the scale in-efficient of the branch companies in three northeastern provinces is mainly caused by the small scales;(4)using three-stage DEA to estimated the efficiency is higher than using SFA. Overall, this paper tinks three-stage DEA method considered of more environment factors, its efficiency measure results are more accurate than SFA method.

Keywords: DEA, SFA, Logistics Efficiency.

1 Dea Model Introduction

Assume presence n different areas as a comprehensive evaluation of decision making units, each decision-making unit has m types of "input" indicators, and s kinds of types of "output" indicators, input indicators and output indicators to be evaluated regional logistics the system input - output indicators[1,2,3].

The base model co0072responds to the decision-making unit DMU_{j_0} (BCC) evaluation model can be used following a pair of linear programming to describe the dual problem: If the linear programming problem the optimal solution ω^0, μ^0, μ_0^0 satisfies $V_p = \mu^{0T} Y_{j_0} + \mu_0^0 = 1, DMU_{j_0}$, then Weak DEA(BCC); if satisfy $\omega^0 > 0, \mu^0 > 0$, then DMU_{j_0} is called DEA efficient[4,5].

$$\left\{ \begin{array}{l} \min \left[\theta - \varepsilon \left(\hat{e}^T s^- + e^T s^+ \right) \right] \\ \text{s.t. } \sum_{j=1}^n \lambda_j x_j + s^- = \theta x_0 \\ \sum_{j=1}^n \lambda_j y_j - s^+ = y_0 \\ \lambda_j \geq 0, j=1, 2, \dots, n \\ s^- \geq 0, s^+ \geq 0 \end{array} \right. \quad (1)$$

$$\left\{ \begin{array}{l} \max(\mu^T Y_{j_0} + \mu_0) = V_p \\ \text{s.t. } \omega^T X_j - \mu^T Y_j - \mu_0 \geq 0, j=1, \dots, n \\ \omega^T X_{j_0} = 1 \\ \omega = (\omega_1, \dots, \omega_m)^T \geqq 0, \\ \mu = (\mu_1, \dots, \mu_s)^T \geqq 0 \end{array} \right. \quad (2)$$

$X_j = (x_{1j}, x_{2j}, \dots, x_{mj})^T > 0$ and $Y_j = (y_{1j}, y_{2j}, \dots, y_{mj})^T > 0$, $j = 1, 2, \dots, n$, the evaluation of the first j_0 , With DMU a non-Archimedean infinitesimal ε , C^2R model can be expressed as [6,7,8]:

$$\left\{ \begin{array}{l} \min \left[\theta - \varepsilon \left(\hat{e}^T s^- + e^T s^+ \right) \right] \\ \text{s.t. } \sum_{j=1}^n \lambda_j x_j + s^- = \theta x_0 \\ \sum_{j=1}^n \lambda_j y_j - s^+ = y_0 \\ \lambda_j \geq 0, j=1, 2, \dots, n \\ s^- \geq 0, s^+ \geq 0 \end{array} \right. \quad (3)$$

2 Index Selection Principle and Determination

Research Logistics of The Northeast Regional, based on access to relevant information on the basis of system analysis and evaluation system to follow certain principles: (1) Science and the main combining principles[9]; (2) quantitative and qualitative principle of combination[10,11]; (3) comparable with the operational phase-principle[12,13]; (4) the predictability of principle[14].

According to the characteristics of the DEA, the DEA method does not require artificial determine the index weight, and the number of indicators not ask for much, only in accordance with the data, including the requirements of the analytical method

will be reflected in many indicators of operating efficiency, select a representative indicator of participation in efficiency measure can. Selected samples in this article are invested with the structure of the regional logistics, output indicators [15,16], such as shown in Table 1. According to the Northeast has the characteristics set logistics input-output system environment variable as follows: (1) per capita GDP (ev1); (2) agricultural percentage of GDP (LEV2); (3) regional market share (ev3).

Table 1. The Northeast logistics efficiency evaluation index system

Input indicators			Output indicators		
x_1 Total assets (million)			y Total profits and taxes (million)		
x_2 Total annual operating cost (million)					
x_3 Number of Employed Persons (number)					

3 Three-Phase Analysis

3.1 The First Phase of Analysis

Select the check information in this article objects to be evaluated for the Northeast 18 regions, to facilitate the calculation and input $DMUi$ for each region, the number of evaluation of this article $DMUi$ 18, respectively, as the decision-making unit in

Table 2. 2006-2010 Manchuria logistics efficiency results

Index	2006			2007			2008			2009			2010				
	DMU	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE	
DMU1	0.15	0.79	0.19		0.29	0.93	0.30	0.43	1		0.43	0.807	1	0.81	0.44	1	0.437
DMU2	0.66	1	0.663	1	1	1	1	1	1	1	1	1	1	1	1	1	
DMU3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
DMU4	1	1	1	1	1	1	1	0.87	1	0.87	0.59	1	0.59	0.39	1	0.39	
DMU5	0.49	0.59	0.83		0.23	0.24	0.93	0.36	0.41	0.90	0.18	0.19	0.93	0.28	0.31	0.95	
DMU6	0.17	0.49	0.36	0.09	1	0.09	0.18	0.73	0.25	0.28	0.60	0.46	0.14	0.58	0.24		
DMU7	0.15	0.88	0.15	0.15	0.85	0.18	0.52	1	0.52	1	0.56	1	0.57	0.42	0.99		
DMU8	0.07	0.70	0.10	0.05	0.99	0.10	0.20	1	0.20	0.20	1	0.20	0.23	1	0.23		
DMU9	0.32	1	0.32	0.26	1	0.26	0.86	0.96	1	0.96	1	0.96	0.12	0.60	0.19		
DMU10	0.30	0.39	0.78	0.51	0.71	0.71	0.53	0.72	0.73	0.78	0.92	0.85	0.77	1	0.77		
DMU11	0.13	0.22	0.62	0.21	0.30	0.99	0.36	0.36	0.99	0.37	0.38	0.97	0.31	0.33	0.94		
DMU12	0.15	0.90	0.17	0.19	0.90	0.20	0.25	0.98	0.25	0.24	0.86	0.28	0.77	0.86	0.90		
DMU13	0.02	0.27	0.07	0.03	0.33	0.09	0.43	0.58	0.73	0.47	0.56	0.81	0.42	0.57	0.74		
DMU14	0.35	0.40	0.87	0.65	0.89	0.73	0.67	0.76	0.88	0.49	0.60	0.82	0.64	0.65	0.99		
DMU15	0.37	1	0.37	0.05	0.66	0.07	1	1	1	1	1	1	1	1	1		
DMU16	0.12	0.36	0.33	0.72	0.77	0.94	0.99	1	0.99	0.73	0.93	0.79	1	1	1		
DMU17	1	1	0.66	0.77	0.85	0.98	0.99	0.99	1	1	1	0.25	0.45	0.55			
DMU18	0.18	0.22	0.82	0.62	1	0.62	0.43	0.49	0.89	0.58	0.70	0.83	0.64	0.75	0.86		
mean	0.37	0.68	0.53	0.43	0.80	0.53	0.61	0.84	0.75	0.62	0.82	0.77	0.55	0.78	0.7		

the evaluation process, namely, $DMU1$, $DMU2$, $DMU18$ the level of efficiency of the logistics system of 18 regional representatives of the level of overall efficiency of the regional logistics system in the Northeast [17,18,19]. Specific samples and data are shown in Table 2.

3.2 The Second Phase of Analysis

Per capita GDP, regional market share and agriculture share of GDP, the proportion of such as environmental explanatory variables into the model of SFA:

$$S_{n,k} = f^n(z_k, \beta^n) + v_{n,k} + u_{n,k}, n = 1, 2, \dots, N, k = 1, 2, \dots, K \quad (4)$$

Among them, $z_k = [z_{1k}, z_{2k}, \dots, z_{pk}]$ ($k = 1, 2, 3$), this set three environment variables can be observed, $f^n(z_k, \beta^n)$ is determined feasible forefront of the relaxation, indicating that the environment variables on the input elements of slack variables $S_{n,k}$ this article take $f^n(z_k, \beta^n) = z_k \beta^n$. $v_{n,k}$ Random factors and assuming $v_{n,k} \sim N(0, \sigma_{v,n}^2)$, $u_{n,k}$, $u_{n,k}$ represents management efficiency factor and assuming $u_{n,k}$ obey the truncated normal distribution, namely $u_{n,k} \sim N^+(\mu^n, \sigma_{u,n}^2)$, assume that $v_{n,k}$ and $u_{n,k}$ is independent of irrelevant. From this we can get the following regression equation [20]:

$$S_{n,k} = \beta_0 + \beta_1^n(ev_1) + \beta_2^n(ev_2) + \beta_3^n(ev_3) + v_{n,k} + u_{n,k}, n = 1, 2, \dots, 10, k = 1, 2, 3 \quad (5)$$

Use the Coelli FRONTIER4.1 software estimated results are shown in Table 3-7.

Table 3. The second stage SFA estimates (2006)

Dependent variable	Workers	the number of relaxation variables	Total	operating costs	Total slack variables	fixed assets of
Independent variables	Coefficient estimates	value of t	Coefficient estimates	value of t	Coefficient estimates	value of t
Constant	630.15**	2.05	33.09***	3.32	-10.82***	-9.77
GDP per capita of the company is located	-0.01**	-2.04	-0.14***	-2.74	1.47*	1.45
Market share	-57.93*	-1.82	925.90***	7.80	34.46**	1.89
Percentage of GDP in agriculture	-10.10	-1.27	96.39***	6.49	4.98***	5.17
σ^2	89.91***	38.89	35.00***	17.54	19.69**	1.97
γ	0.99***	25.8	0.99**	2.10	0.99***	3.18
Log likelihood function		-42.53		-18.70		21.79
LR test of the one-sided error		8.78		16.11		17.58

Table 4. The second stage SFA estimates (2007)

Dependent variable	Workers	the number of relaxation variables	Total	operating	costs	Total fixed assets of slack variables
Independent variables	Coefficient estimates	value of t	Coefficient estimates	value of t	Coefficient estimates	value of t
Constant	348.03**	2.57	12.29***	16.36	36.87***	2.99
GDP per capita of the company is located	-0.01**	-1.76	-0.076**	-2.45	0.05***	3.065
Market share	-40.58*	-1.59	160.27***	6.99	26.12***	14.35
Percentage of GDP in agriculture	-0.54	-1.12	8.07	1.20	10.52***	7.48
σ^2	81.32***	12.76	4.14***	5.74	9.53***	3.64
γ	0.99**	2.50	0.99**	1.84	0.99***	4.18
Log like lihood function		-43.88		-18.78		21.54
LR test of the one-sided error		3.90		9.74		13.25

Table 5. The second stage SFA estimates (2008)

Dependent variable	Workers	the number of relaxation variables	Total	operating	costs	Total fixed assets of slack variables
Independent variables	Coefficient estimates	value of t	Coefficient estimates	value of t	Coefficient estimates	value of t
Constant	-32.82***	-14.45	-69.755***	-5.21	-92.68***	-2.74
GDP per capita of the company is located	-0.001**	-2.67	-0.0544***	-6.59	0.082*	1.53
Market share	-14.77***	-3.12	85.1572***	7.51	93.810***	3.76
Percentage of GDP in agriculture	-1.10*	-1.89	109.1528***	5.87	9.96***	9.80
σ^2	699.99***	97.79	209.7505***	9.75	12.94***	2.54
γ	0.99***	2157.34	0.9998***	9.094	0.99**	1.78
Log like lihood function		-42.72		-31.17		13.77
LR test of the one-sided error		3.82		6.53		5.79

In Table 3-7, $\sigma^2 = \sigma_u^2 + \sigma_v^2$ lead to differences in volatility and logistics management is not valid due to random errors caused by differences in volatility. $\gamma = \sigma_u^2 / \sigma^2$ means that the differences caused by inefficient logistics management changes in the total change in proportion. When the $\gamma \rightarrow 0$ and a certain significant level, significant differences in volatility due to invalid logistics management

Table 6. The second phase of the SFA estimates (2009)

Dependent variable Independent variables	Workers number	the relaxation variables	Total operating costs	Total fixed assets	slack variables	
	Coefficient t estimates	value of t	Coefficient estimates	value of t	Coefficient estimates	value of t
Constant	12.15	1.08	40.83***	3.87	113.42***	3.86
GDP per capita of the company is located	-0.001**	-2.24	-0.06*	-1.66	0.27	1.25
Market share	-1.43***	-4.18	235.52***	13.49	146.86***	3.25
Percentage of GDP in agriculture	-4.76	-1.28	155.10	1.14	52.38**	2.14
σ^2	719.07***	42.89	222.22***	10.65	180.03***	8.75
γ	0.99***	3.046	0.99**	2.09	0.99***	7.26
Log like lihood function		-19.59		-17.65		19.47
LR test of the one-sided error		4.53		6.41		5.09

Table 7. The second stage SFA estimated results (2010)

Dependent variable Independent variables	Workers number	Total operating costs	Total fixed assets	of slack variables
	relaxation variables	relaxation variables	slack variables	
	Coefficient estimates	value of t	Coefficient estimates	value of t
Constant	79.0177*	1.7225	648.8811***	4.6510
GDP per capita of the company is located	-0.0012***	-2.7881	-0.0376*	-1.5474
Market share	-7.4712***	-3.6513	38.3010**	2.1445
Percentage of GDP in agriculture	-2.8337	-1.0764	59.2811**	2.0435
σ^2	418.4966***	34.4241	182.9365***	81.9264
γ	0.9899***	4.0180	0.9992***	7.6372
Log like lihood function		-90.8462		-17.8376
LR test of the one-sided error		3.6382		4.0183
				5.6362

Note: The table γ is the efficiency variation of the ratio of the total variance, and its value significantly, using the Frontier approach is appropriate.

*** Significant level of 1%, ** indicates significant level of 5%, * indicates a significant level of 10%.

can be negligible, that is using the stochastic frontier equation to estimate is inappropriate. When the $\gamma \rightarrow 0$ and a certain significant level, significantly, indicating that differences in impact of the management efficiency in the logistics of production and management process, while the random error of input slack is very small, this time using the stochastic frontier equation to estimate is appropriate.

2006 estimated in this analysis, analysis of other years is similar. Table 3-7 shows the three input variables, the SFA is estimated that the value of γ tends to 1, and each value of γ is significant in 1% or 5% level, indicating invalid sample logistics management. The differences due to random error has little effect on the input factors slack variable, so the use of SFA technology for parameter estimation is appropriate, but this time to consider the elements of value adjustment errors during input.

3.3 Analysis of the Third Stage

In the third stage, according to the adjusted input variables, and adjust the input value of the original output value again on behalf of the into DEAP2.1 software to use the second efficiency calculation based on the input of the VRS model, get the third stage of each DMU true efficiency values, as shown in table 8.

Table 8. 2006-2010 efficiency results in the Manchuria homogeneous environment

Index	2006				2007				2008				2009				2010			
	DMU	TE	PTE	SE TE	PTE	SE														
DMU1	0.61	0.89	0.70	0.59	0.93	0.63	0.91	1	0.91	0.72	1	0.72	0.64	0.64	1					
DMU2	0.89	1	0.89	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
DMU3	1	1	1	0.61	0.75	0.81	0.88	0.88	1	1	1	1	1	1	1	1	1	1	1	
DMU4	1	1	1	1	1	1	1	1	1	0.59	1	0.59	0.55	0.55	1					
DMU5	0.21	1	0.21	0.29	0.30	0.98	0.48	0.96	0.50	0.76	0.83	0.92	0.52	0.97	0.54					
DMU6	0.31	0.53	0.58	0.15	1	0.15	0.20	0.73	0.27	0.31	0.60	0.52	0.46	0.72	0.60					
DMU7	0.15	0.88	0.17	0.15	0.95	0.18	0.69	1	0.69	0.56	1	0.56	0.42	0.42	0.99					
DMU8	0.15	0.70	0.23	0.16	0.98	0.16	0.39	1	0.39	0.21	1	0.21	0.87	0.87	1					
DMU9	0.39	1	0.39	0.44	1	0.44	1	1	1	1	1	1	0.50	0.80	0.63					
DMU10	0.69	0.70	0.99	0.85	0.99	0.86	0.82	0.82	0.99	0.95	0.98	0.97	0.77	0.90	0.86					
DMU11	0.40	0.41	0.96	0.35	0.36	0.97	0.40	0.41	0.97	0.73	1	0.73	0.47	0.96	0.50					
DMU12	0.15	0.90	0.17	0.20	0.90	0.22	0.29	0.98	0.30	0.25	0.86	0.29	0.77	0.96	0.80					
DMU13	0.06	0.27	0.23	0.05	0.33	0.15	0.94	0.96	0.98	0.58	0.69	0.86	0.42	0.79	0.53					
DMU14	0.38	0.41	0.93	0.65	1	0.65	0.95	1	0.95	0.49	0.60	0.82	0.62	0.97	0.64					
DMU15	0.37	1	0.37	0.05	0.66	0.07	0.75	0.78	0.96	1	1	1	1	1	1					
DMU16	0.13	0.36	0.36	0.84	0.89	0.94	0.69	0.70	0.94	0.79	0.98	0.84	1	1	1					
DMU17	1	1	1	0.96	1	0.96	1	1	1	0.62	0.62	0.99	0.37	0.82	0.44					
DMU18	0.39	0.39	0.99	1	1	1	0.51	0.52	0.99	0.62	0.70	0.89	0.64	0.94	0.63					
mean	0.46	0.75	0.62	0.52	0.84	0.62	0.72	0.87	0.82	0.68	0.88	0.77	0.67	0.85	0.79					

Table 8 shows the changes in amplitude of the technical efficiency of the Northeast regional logistics portfolio in 2006-2010 0.459-0.717, in general, showed an increasing trend, and in 2007 reached a maximum value of 0.717. This shows that the current level of logistics management in the Northeast and technical standards still need to be improved and there is much room for improvement. In addition, the logistics and scale efficiency and pure technical efficiency mean in the Northeast in 2006 and 2010 is no more than 0.9, there is much room for improvement, this is mainly due to the logistics

management inefficiency arising. Although some fluctuations from year to year, but overall is still showing an upward trend year by year.

From the combination of technical efficiency point of view, the overall logistics combination of technical efficiency in the Northeast, showing a trend of increasing year by year. The combination of technical efficiency is equal to the decision-making unit decline in the number, a combination of technical efficiency gap among regions be reduced, for the first phase of analysis is a valid amendment.

Change the situation from the point of view, the selected 18 mean pure technical efficiency of sample logistics in 2006-2010 the annual general upward trend year by year more, but the small amplitude fluctuations. The Northeast regional logistics overall level of technology in recent years has improved, and this is consistent with the actual industry trends. Inter-annual fluctuations in its reasons As mentioned above, Since this measure is relative efficiency. From the scale efficiency of scale efficiency of the logistics of the year 2006-2010 the Northeast overall upward trend, but the level of scale efficiency from the ideal level of efficiency is still a big gap. The Northeast all branches of the scale invalid by the small scale due. From the annual distribution of scale efficiency, scale efficiency of 18 samples mean a yearly basis, and the number of effective decision-making scale units increased year by year, indicating that the Northeast region in recent years has made some progress logistics control and optimization of the scale.

4 Comparative Analysis of the SFA and the Three-Stage DEA Results

4.1 The Northeast Regional Logistics Efficiency Empirical Results Contrast

From the empirical method, the SFA to consider non-technical efficiency and random variables on the relative efficiency, three-stage DEA luck and random factors in the logistics management and other business and external factors can be separated in order to improve the traditional DEA model can not solve the measurement the shortcomings of the error. Therefore, the empirical results of the two methods are in line with the actual situation of the Northeast's logistics development and a real and effective.

Mean from each sample, the efficiency estimates obtained in the two methods under the framework each sample logistics efficiency mean there are some differences. Efficiency of each sample using the SFA method and three-stage DEA method estimates the annual mean little difference in the overall use of each sample three-stage DEA method estimates the overall efficiency greater than the results of the use of the SFA estimates, fluctuations in the two methods of calculation results roughly the same magnitude compared in terms of three-stage DEA calculation of the relative efficiency of smaller fluctuations. This three-stage DEA method is not only to consider the random error factors, but also consider the impact of various environmental factors, in the framework of this method, each sample placed in the same competitive environment, the relative efficiency measure. SFA method to consider the impact of random error and inefficiency, but it can not be stripped of environmental factors on the efficiency of the logistics system. There is a big

difference in this article selected a sample of the geographical regions of the total GDP and structure there is a big difference, while the market share of each sample within the region level and facing the competitors, there is a big difference, the competitive environment in each sample differences, such differences are bound to reflect the relative efficiency evaluation results. In comparison, the paper that the measure results of the three-stage DEA method is more objective and scientific.

4.2 SFA and the Pros and Cons of the Three-Stage DEA Measure the Efficiency of Logistics Contrast

SFA and DEA efficiency measure parameters and the typical representative of the non-parametric methods, both in the measure of the efficiency of logistics are applicable. The SFA method taking into account the impact of technical inefficiency and random error, the error in a certain extent, to weaken the efficiency measure, and thus can more accurately measure the levels of efficiency and its variation of the logistics. The actual measurement process, the SFA is the production function-based methodology; it can only measure single output multi-input conditions, levels of efficiency. For the case with multiple output indicators, generally can be used principal component analysis to extract the common factor of the number of output indicators as output indicators, but this would produce a certain measurement error. Less output indicators, this study does not have the use of principal component analysis of the conditions, so only the main output value of the pointer that the total annual profit and tax as a production function regression analysis the dependent variable. In addition, the SFA method to measure the process of the actual production process through the introduction of the logistics of the production function analog samples, and then calculate the production frontier. Because the output of each sample logistics in the measure of technical efficiency, each sample has its own production frontier. Therefore, the SFA in the measure of the level of efficiency of logistics is the actual output of the sample and the sample has a maximum output compared to the same level of inputs to derive the level of efficiency.

Compared with SFA, DEA method is more suitable to handle multi-input, multi-output efficiency of the system level, for multiple output indicators do not need to be processed directly substituted into the planning model calculation. Measure results, in the same decision-making unit and the sample data under the conditions of the SFA and DEA measured by the difference between small, the trend of efficiency over time should be broadly similar. Because the three-stage DEA to consider the SFA to consider the technical inefficiency and random error, and also takes into account the influence of environmental differences on the level of efficiency of decision making units, the measure results relative SFA more robust. Decision-making unit in the advantage of the competitive environment, the final results of the three-stage DEA measure is generally lower than the SFA and measure results.

Contrary, the decision-making unit at a disadvantage the competitive environment, use of a three-stage DEA method to measure efficiency value higher than the SFA. Overall, this paper considers three-stage DEA method is more to consider environmental factors, the efficiency measure more accurate results compared to the SFA method.

References

1. Charnes, A., Cooper, W.W., Rhodes: Measuring the efficiency of decision making units. *European Journal of Operational Research* 6(2), 429–444 (1978)
2. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30(9), 1078–1092 (1984)
3. Battese, G.E., Corra, C.S.: Estimation of a Production Frontier Model with Application to the Pastoral Zone of Eastern Australia. *Australian Journal of Agricultural Economics* 21(3), 169–179 (1997)
4. Yang, F., Bi, G.: Data packets based on competitive decision-making unit Envelopment analysis model. *Journal of Systems Management* 18(3), 332–337 (2009)
5. Lu, W.: Based on DEA evaluation model and its input and output indicators to determine. *Packaging Engineering* 6(12), 119–123 (2005)
6. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30(9), 1078–1092 (1984)
7. Battese, G.E., Corra, C.S.: Estimation of a Production Frontier Model with Application to the Pastoral Zone of Eastern Australia. *Australian Journal of Agricultural Economics* 21(3), 169–179 (1997)
8. Battese, C.E., Coelli, T.J.: A Model for Technical Inefficiency Effects in a Stochastic Production Frontier for Panel Data. *Empirical Economics* (20), 325–332 (1995)
9. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science* 30(9), 1078–1092 (1984)
10. Banker, R.D., Morey, R.C.: The use of categorical variables in data envelopment analysis. *Management Science* 32(12), 1613–1627 (1986)
11. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. *European Journal of Operational Research* (2), 429–444 (1978)
12. Christensen, L.R., Greene, W.H.: Economics of Scale in U.S. Electric Power Generation. *The Journal of Political Economy* (84) (1976)
13. Charnes, A., Cooper, W.W., Rhodes: Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research* 6(2), 429–444 (1978)
14. Luo, D., Xiao, D.: Data envelopment analysis DEA in the power industry. *Theory and Practice of Systems Engineering* (4), 60–66 (1996)
15. Yang, Y.: Economic quantitative analysis method. Jilin Science and Technology Press (2001)
16. Chai, X.: Chain operator performance evaluation indicators and methods. *Business Research* (5), 82–84 (2000)
17. Tu, J., Wu, G.: Based on DEA Tobit two-step method of regional agricultural innovation system evaluation and analysis. *Quantitative & Technical Economics* (4), 136–145 (2006)
18. Huang, X., Yu, D., Willow: CPI Northeast company commercial bank X-efficiency based on empirical analysis of the DEA three-stage model. *Quantitative & Technical Economics* (7), 80–91 (2008)
19. Joe, Z., Wu, G.: DEA theory, methods and applications. Science Press, Beijing (1996)
20. Wei, Q.: Data envelopment analysis. Science Press, Beijing (2004)

Idols Supply Pattern in Confident Reestablishment of National Culture

Shanguang Lv, Huifen Li, and Xiuxia Wang

Shandong Institute of Commerce and Technology the Teaching Section of Ideological and Political Lessons Shandong Jinan 250100

Abstract. The cost of information communication is so cheap in the information age that makes idols supply with a long history of moral education mode change quickly from the traditional and single power oriented pattern to the structure that combined power leading mode and social leading mode. The idols supply pattern dominated by power leading mode and the other pattern effect meet the unprecedented challenges and huge interference brought by the society oriented idol supply pattern. And they two compete for the limited theory and leading space which relate to their survival in the same technology platform. Under the new situation, our cultural field is undergoing extensive and profound changes. And promoting cultural development and prosperity possesses many favorable conditions, and faces a series of new conditions and problems as well. So we need to perfect the idol supply pattern from multi-aspect and multi-angle, so that we can rebuild our national cultural self-confidence.

Keywords: Idol, Idol Supply, National, Cultural Self-Confidence.

1 Introduction

According to the China Youth Daily Social Investigation Center recently conducted a project entitled "Chinese culture class awards Why the lack of influence" of thousands of people online survey showed that 37.6% of respondents of various cultural awards "concern". Recently, the Social Investigation Center of China Youth Daily conducted a thousand of people online survey entitled "Why Chinese cultural awards lack of influence?" And it showed that 37.6% of respondents concerned about our present cultural awards, 32.8% of the people didn't concern any more. During the survey, 81.4% of respondents believed that China had few cultural awards that with credibility in the present. And 22.3% of the people thought we have very little. Only 3.7% of the respondents thought we have a lot, and 13.4% thought it is OK. Among these respondents, post-80s occupide 40.1% and post-70s took up 40.2%. What's more, "Mao Dun Literature Award" has been questioned as President Prize for Literature and "Huabiao film Award" is ridiculed as the commendation party for filmmakers and the like. Chinese cultural awards are suffering the crisis of public trust on some level. The result of survey has been alarming from some side that we are facing the crisis of public trust.

On October 18th, 2011, the Sixth Plenary Session of the 17th session of the Central Committee of Communist Party of China passed the document, the decision to deepen cultural restructuring to promote the development and prosperity of socialist culture and a number of major issues. The "Decision" pointed that the Culture is the lifeblood of the nation and is the spiritual home of the people. But with the development of society, our cultural field is undergoing extensive and profound changes. And promoting cultural development and prosperity possesses many favorable conditions, and faces a series of new conditions and problems as well. Chinese cultural awards are suffering the crisis of public trust on some level. In fact, it is the highlights of these new situations and problems. What's more, the result of China Youth Daily Social Investigation Center's survey as well as CPC Central Committee's decision becomes the practical and theoretical starting point for the authors when they writing and discussing.

2 The Rise of "Network Culture" of Post-industrial Society Reflects the Crisis of National Cultural Trust

American sociologist D Bell forecasts in his book, "The advent of post-industrial society", that the social structure of post-industrialization will eventually lead the United States, Japan, the Soviet Union and Western Europe to appear a new form of society in the 21st century. Moreover, the rise and development of "network culture" are the main characteristics of this new form of society. The low-cost, speediness, standardized difference and extensive coverage of "network culture" will inevitably lead the whole society to be "imetusous" "emotional" and "utilitarian". And all these happen to be the psychosocial overall performance of this new social form.

2.1 Fickleness Replaces Reality

When the commercial value becomes the society's most important value orientation and when flourishes effect of gamble and net celebrity become popular overnight, which can get huge economic interests, the whole society will form an overwhelming psychological inertia. And it makes all the people's psychology of the society tend to be impetuous. Dreaming something with no pains and flourishes makes people prefer to be opportunistic and instantly successful rather than struggled. Besides, the purchase of lottery tickets, the bankrupt of a single family, the high illegal interest rates as well as a wider range of social unrest are all the adverse social consequences brought by the social mood of "impetuous".

2.2 Moodiness Displaces Rational Thinking

The glorious of money makes people lose their compassionate natures by replacing a favor of extreme cold and the supremacy of money. It fully disclosures that social psychology has been turned from rationality to emotion when taking driving luxury cars, treating lives as worthless and showing domineering family into consideration. A lot of individual behaviors and emotions are added in social life, which will inevitably

lead to a social psychological impulse to replace the rational framework for thinking built up by the hard work of mainstream culture.

2.3 Utilitarian Replaces the Public Interest

"Valuing Loyalty" advocated by the fine traditional Chinese culture has been replaced by naked "utilitarian" in the information age. "Do unto others, do not impose on others" has long been mercilessly repressed by the economic interests of the center lever of market economy. The "waste oil" and "toxic milk" are all reflections of the "Wang Yi materialistic" mentality with which individuals are driven by economic interests under the conditions of market economy, regarding the public interest.

Network, one of the most efficient ways of information dissemination plays a huge role in the erosion of the foundation of national culture in modern society which is full of "impetuous", "emotional" and "utilitarian". The fast-food type of network culture caters to the tastes of several generations of people who grew up with the network technology with its ultra-fast, low-cost and diversity, just like Taiwan campus songs which can be famous over one night in 80s. Moreover, the rise of "network culture" leads many "Network celebrities" to become "teen idols" of the younger generation quickly. Fishes and dragons jumbled together in the inside of community-led idol supply system that covered many netizens' physiological function and value criteria. The allatonce and speed of idol supply weaken connotation of people's hard work and imitated learning, which becomes another "Idol Casino" of cultural dimension in the information age. And this emotions and thinking inertia of "idol Casino" further deepen" impetuous ", " emotional "and" utilitarian " in modern society and the crisis of national cultural trust.

3 The Second: The Revitalization of the National Culture Needs Our Reflection on the "Idol Supply Mode"

In the information age, the unprecedeted emptiness of soul and the eager to find the spiritual comfort and sustenance of the spiritual homeland has changed from the individual's mental illness to the social feature ,with the main carrier called Internet culture" .In fact, there are many reasons which lead to the crisis of the national culture, but the generalization of the "supply modes of the idol " is the most important one, with the background of the "net culture " ,the arrival of the modern society do not led to the downfall of the idols because of the rise of the various individual philosophy, even we can say, due to the formation of a strong state of the mass media and the fall of public spirit, more and more heroes and wise men needed, as well as the excellent men who can prove the significance of human beings, we can say that modern society and daily life need idols "[4] In the information age, the low cost of communication makes the method of the idol production go away from the traditional route and can be completed by self-help, which makes it possible in theory that' everyone can be idols'. Some poor idols also may be worshiped and followed .All things indicate that we have entered a 'pan-idol' era. ", [5] In this historical background, the supply of idol, the moral education which has a long history has changed by traditional single-power-led to a

power-oriented social oriented binary coexistence of structure rapidly. The supply mode of idols with power and the effects of pattern after that encountered unprecedented great challenges from society under the guidance of the supply modes of idols, both of them compete on the same technology platform for the limited leading spirit space.

Traditional single-power-led "idol supply mode" has been challenged powerfully by the modes of "social leading mode". The main idol type of mainstream values, containing the value of content has been covered by the non-mainstream values, national culture became disorder, which is the major cause to the crisis of confidence in the national culture. National culture, the deepest elements as the shoulder of a society and the country's future and fate, will breakthrough with the reflection and build of "idol supply mode".

3.1 The "Idol Supply Mode" Should Continue to Develop with the Expansion of the Vision in This Era

Relying on science and technology, the development of this era promotes the progress of human society .With the development of the times, as a heritage of excellent national carrier of culture, the idol supply mode" should also constantly change perspective and expand sight. Under the planned the development of this era Under the planned economy, the value of people's social life has been unitary and patterns, and "willing to benefit others instead of themselves "is not only the expression of people's aspirations but also the basic norms of everyday but also the basic norms of everyday behavior, occasionally rebellious behavior that deviates from the mainstream value will would be the object of that era . Under market economy conditions, the acquisition of economic benefits has become a basic value orientation of people's behavior, while their adequate exchange, contrary to the economics the economics law is chasing the object of economic behavior. From the sale of shoddy toxic food drug production which is disregard of human life and the safety health to killings for money even daughter violence, all these extreme social consequences brought from the blind spots of "supply mode of idols". Adapt to the times and changes, the rebuilding of the" supply model of Idol" has constantly opened up from different perspectives such as distance, depth, and height.

3.2 The Construct of "Idol Supply Mode" Should Be Begun from the National Long-Term and Fundamental Development

We explore the construct of "idol supply mode" just for the benefit of our nation's long-term and fundamental interests, and it indeed should be begun from the national long-term and fundamental development. Firstly, our "Idol supply mode" must be rooted in the outstanding traditional culture and draw the ingredients that full of vitality from it, which can make it full of a timeless value of life; Secondly how we choose or build "idol supply mode" must take the long-term and fundamental interests of our nation into account and then make rational choices on a solid ethical and legal basis; Thirdly ,absorb all the other nations' outstanding cultures on the premise that it shows full respect for our national long-term and fundamental interests. And only in that case, the choice of "idol supply mode" can gain wider recognition and support.

3.3 The Construction of “Idol Supply Mode” Should Closely Be Combined With the Development of Modern Science and Technology

The development of modern science and technology has increasing influence on the other activities of human beings. The selection and construction of “Idol supply mode” in human spirit field also cannot be divided from the support and restrain of modern science and technology.

On the one hand, the selection and construction of “Idol supply mode” must take fully advantage of the power and technical support of modern science and technology, defusing the technical influence of modern science and technology which are widely controlled by people; on the other hand, we should hold a favorable predict and control of modern science and technology and its future trend, and always grasp the opportunities of science and technology development.

4 To Prove "Idol Supply Mode" in Order to Rebuild National Cultural Self-confidence

"The value of diverse social forms creates a pluralistic of society idols Social idol of itself has already been the leaders of the social practice, the communicators of Social civilization and the practitioners of social morality". Improving “idol supply mode and rebuilding the national cultural self-confidence become the theme of the times and historical tasks. Besides, Central Committee of the CCP points in< about deepening the reform of the cultural system, promoting socialist cultural development and prosperity of certain major issue decision> that we should develop the democratic, scientific and popular socialist culture which is facing modernization, facing the world and facing future, develop a high degree of cultural awareness and cultural self-confidence, improve the ethical quality of the whole nation, increase the soft power of the national culture, promote the Chinese culture and make efforts to built powerful country with Socialist culture."This ambitious goal provides us the driving force to improve the “idol supply mode in order to rebuild cultural confidence. When considering the "Idol supply mode" building issues, we should look at the following aspects in order to achieve the grand goal of building of the socialist culture and powerful country.

4.1 Strengthen the Construct-Oriented of “Idol Supply Mode” of Mainstream Media

“The mass medium is such things as newspaper, magazine, broadcast radio, TV and Internet. They play an extremely important role in guiding the public opinion of today's society and they are the important communication carrier and survival foundation for idol culture.” The rapid development of information society has made the mainstream media become integrated media groups of the party newspaper of all levels, TV station, the government mainstream web, into which the past Two Newspapers and One magazine. But the increasing channel of mainstream is not in proportion to the mainstream media public opinion direction effect. The blundering, emotional, and utilitarian mental representation of the times spread over the huge amount of information transmitted by the mainstream media. Indeed, the mass media of modern society must take both informed rights and feelings of people into consideration,

however, the utility effect of “attention economy” would exclude the media, weakening the functions of its positive guidance of Public Opinion, even making it go over to its opposite site. Therefore, the positive mainstream-oriented role of mainstream media in building “idol supply model” definitely cannot be weakened at any time, any circumstances.

4.2 Master the Direction of “Idol Supply Model” of Network Media

In terms of the speed, cost and influenced sphere of information spreading, the Network Media has its own advantages that the other medium don’t have, which exactly make it hard for the “Idol supply model” admired by the Network Media to master the trend. With the encouragement of communism agitation and the promotion of mass culture and business culture, civilian idols and practical idols----all kinds of heroes of new wealth are starting to impact traditional example images seriously, the new models and new idols aimed of realizing their own values have stepped into the stage” It could be popular to be an Internet star all around the world overnight, it could be known for notoriety because of some accident. We can experience the power of online media in various boffo incidents, from the “Ugly sister Furong Runner Fan” to “Coat brother” Sister Furong”. We must have new thinking patterns and new measures to adjust to the effective control actions and technologies to master the direction of “Idol supply model” of net media under new situation, which wouldn’t lose the battle field of the guidance of public opinion.

4.3 Implementing the Effort of the Ideological and Political Education “Idol Supply Model”

Any society of the ideological and political education work bearing the society’s important load, a mainstream social values as well as overall social value orientation to a large extent depends on the success or failure of the social ideological and political education work. It goes without saying that the ideological and political education is particularly important to the fundamental role of the“idol supply model”. From April 1980 the Ministry of Education, the Central Committee jointly issued the "views on strengthening the students' ideological and political work in colleges and universities, to 2004 the CPC Central Committee and State Council issued the" Opinions "on Further Strengthening and Improving Ideological and Political Education, twenty years of thinking about the practical course, fully illustrated the important role of the ideological and political education as well as party and state attention, it is also all levels of education institutions of ideological and political education work productively, to ensure us the cause of eternal youth and success or someone. For the contemporary college students, idolatry has no previous generation of the sense of the solemn and sacred sense, the Ideological and political education should be combined with a model education to correct and timely guidance on the idolatry phenomenon of university students. Under the new situation, with the diversification of means of information dissemination and efficient, it is important to further strengthen the ideological and political education and reinforce the role of the foundation built in the areas of social life "Idol supply model".

5 Forming the Building Mechanisms of “Idol Supply Mode” in Whole Society

The construction of “Idol supply mode” of a society needs the entire power of every aspect in society. Both of the building guiding role of the main media’s “Idol supply mode” and building basic role of “Idol supply mode” of ideological and political education should form a Social whole mechanism under the guidance of the government, and every process is just an indispensable, interdependent and synergistic part of the system, which requires the whole society to form a macro coordination and operating mechanisms of “Idol supply mode”. The social dominated “Idol supply mode” which characterized by network culture mustn’t be let aside. We should hold the initiative right of the construction of “Idol supply mode” on the basis of guide, specifications and persuasion.

References

1. The Beijing News survey. How cultural awards create influence, Beijing, 4th (B03) (February 2012)
2. Central Committee of the CCP about deepening the reform of the cultural system, promoting socialist cultural development and prosperity of certain major issue decision. In the International (2011)
3. Central Committee of the CCP about deepening the reform of the cultural system, promoting socialist cultural development and prosperity of certain major issue decision. In the International (2011)
4. Jiao, C.: Some analysis of Ideological and political education idol supply mode. Study of Ideological Education, 1st (2012)
5. The current viewpoint. We should be able to find ourselves in the era of "Pan Idol" and "anti-idol". Campus psychology, 3d (2006)
6. Jia, R.: The social idol's moral examine of value of diversity background. Study of Ideological Education, 2d (2012)
7. Central Committee of the CCP about deepening the reform of the cultural system, promoting socialist cultural development and prosperity of certain major issue decision. In the International (2011)
8. Peng, X., Sun, Y.: talking about the contemporary college students' idol worship. Liaoning University of Technology (Social Science), 10th (2011)
9. Wen, H.: Analytic to young idolatry phenomenon changes. China Youth Study, 8th (2008)
10. Guo, R.: Students idolatry phenomena and Education Strategy. Shanxi Agricultural University (Social Science), 10th (2011)

Research of Popularizing Solar Heating in Countryside

Junjun Wu, Bin Xia and Mingchuan Meng

Qing Gong College of HeBei United University, Tang Shan, 066003, China
junjun-yd@163.com

Abstract. With the development of rural economy, energy demand is also growing, in order to solve the problem of rural energy, ecological villages and towns should establish a point of view, focusing on the development of renewable energy. Solar energy is very rich in renewable energy, pollution-free environment. Solar energy is more and more important to people, so and the technology of solar energy. For the study to the rural areas of North China, Paper analyzing the feasibility of popularizing solar heating in countryside from the advantage of solar heating, the advantage of solar heating in countryside and the system of solar heating, At the same time, we study the key technologies of solar heating, such as the choice of collector, the choice of the thermal storage tank volume. It shows that there is a great value and significance promoting solar heating in countryside, especially solar radiant floor heating. Popularizing the use of solar heating can reduce greenhouse gas emissions and pollutant emissions of conventional energy use, minimizing burden of rural heating, it is clean, health, safety, meeting the objective requirements of the new rural construction. In the new rural construction, solar heating has better social benefits and promotion prospects; there is a positive role in promoting China's economic, environmental and energy construction.

Keywords: Solar Heating, Solar Collector, Popularizing.

1 Introduction

With the development of the national economy, energy demand is increasing, and energy using is nervous, while large scale using of conventional energy is bound to adversely affect the environment. Solar energy as a renewable energy, inexhaustible, without increasing the environmental load, will become an important part of future energy mix. China is one of the solar energy resource-rich countries, the annual total radiation is about 3300-8300 MJ/(m².a), annual sunshine hours of 2/3 area is greater than 2000h, each year the solar radiation the land receive is equivalent to 2.4 thousand billions tons of standard coal, with good conditions for utilization of solar energy. In building energy consumption, domestic hot water and heating accounted for a considerable proportion in energy consumption, using solar energy to meet this base-grade requires such as domestic hot water and heating can bring huge energy efficiency, solar heating technology is becoming increasingly important to the people [1-3].

2 Analysis of the Feasibility of Promoting Solar Heating in Rural Areas

2.1 The Advantages of Solar Heating

Theoretical research and practice shows that the solar heating program suitable for rural living conditions. solar heating has been implemented in the rural areas of China's different regions, like Chang ping Buddha Deer Village, the Ping gu Village and glass Taiwan Village in new rural construction demonstration in Beijing. The technology is reasonable, operation is simple, it get the praise of villagers. In the areas that not have implementing solar heating in winter, heating is heated by adobe stove and coal stoves which are not very hygienic methods of heating, adobe stove and coal stove will have a lot of flue gas, slowly blacken walls of the indoor, especially coal stove, if the ventilation system placement is not good, but also lead to gas poisoning, while solar energy is a clean, safe, inexhaustible energy, with the improvement of energy technologies in China, the promotion of solar heating has great value and significance.

North China rural areas need the heating time from mid-November in winter to the second year in mid-march, Horizontal solar direct radiation in heating period and the total amount of radiation is given in accordance with meteorological statistics of the typical year, calculate the amount of solar radiation on an inclined plane of about 49 degrees between North China and the horizontal angle with horizontal average daily solar radiation values.

According to formula (1) the calculated results for \bar{H}_{Th} is [4]:

$$\bar{H}_{Th} = \bar{H}_{bh} \bar{R}_{bh} + \bar{H}_{dh} \left(\frac{1+\cos S}{2} \right) + \bar{H}_h \left(\frac{1+\cos S}{2} \right) \rho \quad (1)$$

$$\bar{R}_{bh} = \frac{\frac{\pi}{12} \sin(\phi - S) \sin \delta + \cos(\phi - S) \cos \delta \sin 15^\circ}{\frac{\pi}{12} \sin \phi \sin \delta + \cos \phi \cos \delta \sin 15^\circ} \quad (2)$$

In the formula \bar{H}_h , \bar{H}_{bh} , \bar{H}_{dh} is the horizontal total global solar radiation, direct radiation and diffuse radiation respectively. ρ Is surface reflectivity, \bar{R}_{bh} is the ratio of the inclined plane and horizontal beam radiation, ϕ is geographic latitude, S is the angle between the horizontal plane and inclined plane, δ is the local declination angle.

2.2 Advantages in Test-Practice of Solar Heating in Rural Areas

Relatively high-rise buildings of the cities, rural solar heating has certain advantages. Rural housing is mostly hilly scattered arrangement and scattered, it is efficient to receive the energy of solar radiation, which provide for external feasibility of the operation of solar collectors. Moreover, the building density of rural areas is small. residential height difference of rural building is very small, unlike the tall buildings of

the city blocking the sun each other .the rural housing has a great advantage in that regard, the advantages of building orientation is to use sun geothermal resources Sufficiently and reasonably. It requires that solar collector and the building itself as much as possible to receive solar radiation in the time of heating. But in the summer in order to avoid indoor overheating, you should minimize direct sunlight to room. From the law of solar activity at high latitudes, the south can meet the requirements toward. In addition, the selection of the south side door also has a close relationship with seasonal wind. the cold northwest wind is popular in typical heating area in winter, while the prevalence of summer is warm and humid southeasterly winds, which also determines the basic pattern of the building should be facing south[5-7]

2.3 Solar Heating System

The schematic of solar heating system is shown in Figure 1.

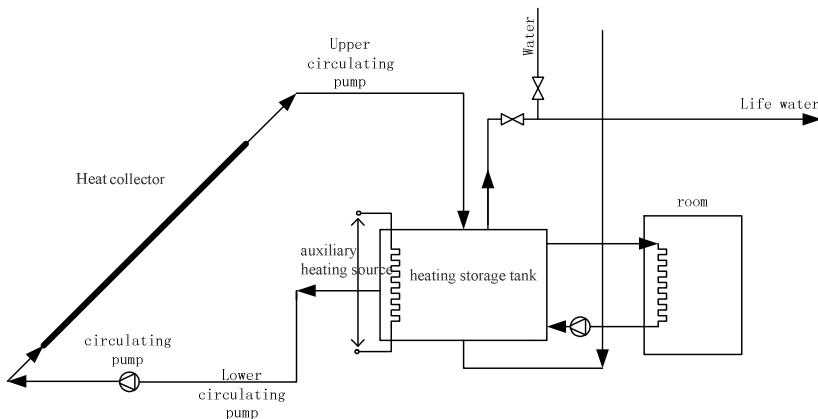


Fig. 1. The schematic of solar heating system

(1) Solar collector

An important component affecting the performance of solar heating systems is the solar collector. Therefore the key to the promotion of engineering applications is the quality of the product of solar collector. Solar collectors mainly is in the form of two kinds of flat-plate collector and evacuated tube collectors, vacuum tube collectors hold domestic solar water heater market, but in solar heating system, flat-plate collector because of their material, structure, the advantages of working life and with construction combined with the appearance is more and more popular to people, and most of the solar heating works combine with building form, so it has the high demanding in performance of combining with construction, the low failure rate, life of products compared to the all-glass vacuum tube and heating pipe or vacuum tube solar collector, flat-plate collector has the characteristic of simple structure, compressive strength, resistance to external shocks, resistance to hot and cold shocks, low failure rate, long life and easy to combine with the construction. Vacuum tube and heat pipe

collector problem of relatively high failure rate ,short life, poor performance after combining with building and other issues with the. As mentioned above. Compared to parallel vacuum tube collectors, despite the inferiority of the insulation performance of the flat-plate collector, but its effective lighting is much larger than the vacuum tube collector and the total useful area of solar energy is higher than the vacuum tube collectors, so we use flat-plate collector.

The formula for calculate the area of solar collectors is [8]

$$A = Qf[H \times \eta_{cd}(1 - \eta_L)] \quad (3)$$

In the formula, A is light area of solar collector, Q is the heat load of the building heating, F is the solar guaranteed rate, H is radiation of collector solar MJ/(m².d), η_{cd} is average collection efficiency of collector, η_L is heat loss and it is 0.2.

(2) Solar heating storage tank

Solar heating system, hot water systems and heating systems have different request, the lower the operating temperature of the solar collector system, the higher the thermal efficiency, So system designing should minimize the operating temperature of the solar collector system. The low temperature floor heating system is suitable for solar heating, water temperature at 40°C; domestic hot water supply temperature at 50°C~60°C. To satisfy different water temperature requirements, solar heating systems adopt generally vertical stratification tank. The works of vertically stratified water tank is the use of the density of water at different temperatures to achieve different partition of temperature in the same tank; low temperature water is in the bottom of the tank, hot water in the upper of the tank, and not mixing with each other. At the bottom of the stratified tank we lay the heat exchanger connected to solar collectors; the water temperature in the middle is suitable for heating which is connected with the heating system; the upper water temperature is the highest, layouting heat exchange of the domestic hot water.

Regenerative collector efficiency and the heating load has a relationship with solar guaranteed rate, the specific heat storage can be expressed as:

$$Q = \int_{T_1}^{T_2} \rho V c_p dT \quad (4)$$

In the formula, T_1 and T_2 is the initial temperature and heat storage temperature of the thermal storage medium respectively, for heating and water systems, usually using water as a heat storage medium, the specific heat capacity and density of water can be seen as a constant, so the volume of thermal storage tank can be [9]:

$$V = \frac{Q}{\rho c_p \Delta T} \quad (5)$$

The size of the thermal storage tank capacity is the system parameters that need to optimize. Small storage tank capacity bring high efficiency in the use, but if the temperature rises too fast ,it will reduce the efficiency in the use of the collector, large thermal storage is not full use, and increase the heat dissipation of the whole system, so we need to weigh carefully.

(3) Auxiliary heating source

The water supplying characteristics of solar heating system for residential heating with hot water and solar water heating systems providing hot water for residential are different, the domestic hot water does not require a continuous supply, but heating and hot water must be continuous supply, and reliable. Solar radiation is restricted by natural conditions such as day, night, season, latitude hair and altitude, and it is influenced by weather and other random factors, there is a certain amount of instability in the solar heating system, you must set the auxiliary heat source. In the northern region, the auxiliary heat source can be electric furnace, camp stove, and so on [10-11].

3 Analysis of the Economic Benefits of Solar Heating

Relative to coal, the heat of coal is 20.934MJ/kg, according to the current market, the price of coal is about 1.5 ¥/kg. If efficiency is 65%, the hot price is 0.1102 \$/MJ; And heat value of natural gas is 35.17MJ/ m³, relative to natural gas, unit price is 2.3 ¥/ m³ in calculation .if the efficiency of the system take 85%, the hot price is .0769 \$/MJ; relative to electric heating, if electricity price is 0.6\$/kwh in calculations, the efficiency of the system is 90%, the heat price is 0.185 ¥/MJ, the annual cost savings and payback period is shown in table 1. From the recovery period of investment, solar heating replace electricity and other cost-effective energy, static investment recovery period is about six or seven years, to the alternative energy at a lower price such as coal, natural gas, static investment payback period is longer than the life of the system, mainly because The actual utilization of solar energy systems in non-heating season is not high, the return on investment is mainly reflected in the social aspects. Therefore, the promotion of solar heating technology in the new rural construction also requires government subsidies.

Table 1. The annual cost savings and payback period

area of Solar collector (m ²)	10	15	20	25
System cost + maintenance cost (million)	3	3.5	4.0	4.5
Effective heat gain (MJ)	18251	21432	25303	31072
payback period instead of coal (years)	12	14	16	18
payback period instead of electric (years)	6	7	8	9

4 Conclusion

Studies have shown that promoting solar heating in rural areas has great value and significance. It can bring economic benefits and environmental benefits, but the return on investment problem may be the factors that hinder the promotion of speed, therefore the promotion of solar heating technology in the new rural construction also need the government to give subsidies. We believe that in the future the environment in rural areas will be much better.

Acknowledgements. Supported by natural science foundation of Qing Gong College of HeBei United University under Grant No. qz201009.

References

1. Cang, L., Li, J., Zhou, L., Sang, Y., La, M.: Talking over application and developing of solar heating. *Xi Zang Technology* (10), 8 (2010)
2. Cai, W., Zhang, L., Gong, X., Jie, G.: The analysis of feasibility of solar heating in typical rural house. *NingBo Engineering College Learned Journal* 20(4), 37 (2008)
3. Zhu, D.Z., Lu, C., Liu, J.: The technology and the design of solar heating. *Building Energy and Ventilated Air* 26(2), 52 (2007)
4. Li, N., Lv, J., Yang, H.: Feasibility analysis of different solar space-heating schemes in Tian Jing area. *Acta Energiae Solar Sinica* 30(12) (2009)
5. Zhao, X.-P., Wang, J.-Q., et al.: Analysis on energy consumption of existing building in cold regions. *Xi an University of Architecture & Technology: Natural Science Edition* 42(3), 427–430 (2010)
6. Qian, B.: Solar technology and applications. *Science Press* (August 2010)
7. Hao, J., Zang, P.: Analysis of size planning of solar heating and hot water Combination System. *Building Energy & Environmen.* (9), 30–31 (2009)
8. Xue, Y.-B., Kang, Y.-D.: Economic Analysis on Solar Heating of Rural Residential in Shan dong. *Building Energy Efficiency* (9), 17–18 (2011)
9. Wang, Q., Chen, H.-B., Hao, Y.: Application Discussion on Solar Heating System in Some Areas of Yun nan. *Building Energy Efficiency* (3), 43–44 (2011)
10. Zheng, H.C.: Solar low temperature hot water radiant floor heating. *Solar Energy* (1), 14–15 (2000)
11. Liu, Y., Yu, Y., Kong, D.: Solar fraction of residential building in north-west China. *J. Xian Univ. of Arch. & Tech (Natural Science Edition)* 43(2) (2011)

A Social Benefit Comprehensive Evaluation Model Based on Unascertained Measure Expectation of Science and Technology Information Products

Jianli Chen

Department of Basic, Armed Police Force Engineering University,
School of Science, Shaanxi Xi'an, 710086
zbh62825@163.com

Abstract. By studying the problems of social benefit comprehensive evaluation of science and technology information product, and analyzing the social benefit factors of science and technology information products, it has been established by the index system of social benefit comprehensive evaluation of science and technology information product. By defining unascertained measure model expectations, and combining the method of unascertained measure, it has been given by a kind of compatible comprehensive evaluation method of social benefit of science and technology information product and unascertained measure expectation model. And as an example of “super miniature camera technology”, it has been given by the specific application of evaluation method and model, it was illustrated that the method is scientific and effective.

Keywords: Science and Technology Information Product, Social Benefit, Evaluation, Unascertained Measure.

1 Introduction

Social benefits have broad sense and narrow sense two kinds, that is all the non economic benefits are called social benefits. However the narrow sense social benefits separate the non economic benefits in some aspects of benefits, and together is brought forward by social benefits, the limits is much smaller. The social benefits of this paper studies refer to the broad sense social benefits.

Social benefits is the reflection of the proportionate relationships between social benefits and social costs, but the social interests and social costs are extensively effected by lots and lots of tangible or intangible, controllable or uncontrollable factors from different social sectors, different areas, different regions and different period. Therefore the social benefits measure is very complex, and is difficult to be expressed by the determined formula. Hence analyzing the main factors of social benefits, establishing the evaluation index system, and then progress the comprehensive evaluation, this method has became a common thought of research on social benefits.

On the other hand, many information in social benefits evaluation system is uncertain, that is unascertained, hence using the unascertained measure and measure

expectation model comprehensively evaluate the science and technology information products is a kind of effective method. Since the unascertained measure has been put forward, it has been widely used in many fields, and has achieved good results.

2 Analysis the Social Benefits Factors of Information Technology Products

2.1 The Determination of Evaluation Indexes

Product of information of science and technology social benefits comprehensive evaluation produced the social benefits by the information products in use or after use, such as increasing the accumulation of knowledge[1,2], boosting the level of decision-making, satisfying spiritual needs, promoting the progress of science and technology, promoting the development of society, protecting the ecology environment, increasing social welfare, activating cultural life, promoting academic exchanges, conducting to social stability and so on, in order to comprehensively evaluate and analyze the non economic factors that brought the total social benefits, through the discussions with experts, to determine the influence of main factors of social benefits. Evaluation factors were chosen for the products of specific information and its application dependent, different information products can choose different factor of evaluation. Such as the main factors of social benefits evaluation of "super miniature camera technology" includes: f_1 : promote the progress of science and technology, f_2 : promote economic development, f_3 : strengthen social stability, f_4 : save social resources, f_5 : reduce the burden of consumption, f_6 : boost decision-making standard, f_7 : satisfy security needs, f_8 : active cultural life, f_9 : fill gaps of the market, f_{10} : improve nation's reputation. The above ten aspects constitute the social benefits evaluation factor set $F: F = \{f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9, f_{10}\}$.

2.2 Evaluation Level of Classification

Through discussions with experts to determine the evaluation factor's evaluation scale set $V: V = \{V_1, V_2, \dots, V_n\}$ evaluation of scale set is also called comment set [3,4], it is the possibility that the valuator makes the social benefits of information technology products as the total judgment level. As a super miniature camera technology information products can be divided into social classes V_1 very poor, V_2 less poor, V_3 normal, V_4 better and V_5 best.

The above mentioned five grade of evaluation elements constitute class set: $V = \{V_1, V_2, V_3, V_4, V_5\}$.

According to the evaluation index of the standard, it can specify the definition for each evaluation grade concretely.

3 The Unascertained Measure Expectation Model of Social Benefits

Suppose some science and technology information product that has m evaluation [3,4, 5]index I_1, I_2, \dots, I_m , then the evaluation index space is $I = \{I_1, I_2, \dots, I_m\}$. Thus, x can be expressed as m dimensional vector $x = \{x_1, x_2, \dots, x_m\}$, where x_i means the evaluation of the object of evaluation index I_i 's evaluation value. For every x_i , that has P evaluation grade, C_1, C_2, \dots, C_p , and then the evaluation of space is $U = \{C_1, C_2, \dots, C_p\}$, suppose C_k means social benefits, grade $k+1$ is higher than grade k , denoted by $C_{k+1} > C_k$. If $\{C_1, C_2, \dots, C_p\}$ satisfy $C_p > C_{p-1} > \dots > C_1$, then we can call $\{C_1, C_2, \dots, C_p\}$ is an ordered partition of evaluation of space V.

3.1 The Index of Unascertained Measure

Let $\mu_{ik} = \mu(x_i \in C_k)$ show the evaluating values of x_i that is belong to the k th evaluation grade C_k 's extent, request μ to meet the condition:

$$(1) 0 \leq \mu(x_i \in C_k) \leq 1, (2) \mu(x_i \in U) = 1, (3) \mu(x_i \in \bigcup_{i=1}^k C_i) = \sum_{i=1}^k \mu(x_i \in C_i).$$

Where $i = 1, 2, \dots, n, k = 1, 2, \dots, p$, (2) is called the normalizing, (3) is called additivity, the μ that meet the condition (1), (2), (3), is called unascertained measure[5].

3.2 The Weights of Evaluation Indexes

Use the w_j to express the evaluating indicator I_j compare with other indicators it has a relative important degree, request w_i to meet the condition: $0 \leq w_j \leq 1, \sum_{j=1}^m w_j = 1$

Let w_j be the weight of I_j , calling the vector $w = (w_1, w_2, \dots, w_m)$ as the index weight factor, in the unascertained comprehensive evaluation system, the index weight is very important. According to the experience of experts to determine the weights of important degree of the factors, it shows the important degree or contribution of the factors on social benefits target, according to certain rules given by experts of each index "score", and use "statistical scoring" method to determine the index weight vector[5,6].

3.3 Comprehensive Measure Evaluation Matrix

Let $\mu_{ik} = \mu(x_i \in C_k)$ express technological information products x_i belong to the k -th evaluation grade C_k degree, therefore $\mu_k = \sum_{j=1}^m w_j \mu_{jk}, k = 1, 2, \dots, p$. Since $0 \leq \mu_k \leq 1$

and $\sum_{k=1}^p \mu_k = 1$ so μ_k is unascertained measure, calling $(\mu_1, \mu_2, \dots, \mu_k)$ as the comprehensive measure evaluation vector [6-8]. Calling the matrix

$$(\mu_{ik})_{m \times p} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{m1} & \mu_{m2} & \cdots & \mu_{mp} \end{bmatrix}$$

as the comprehensive measure matrix.

3.4 Unascertained Measure Expectation Model

Definition Let $[[a, b], \phi(x)]$ be an unascertained number, $\phi(x)$ is an unascertained measure, calling $E = \int_a^b x\phi(x)dx$ as unascertained measure expectation[9,10].

Let $\phi(c_i) = \mu_i = \sum_{j=1}^m w_j \mu_{ji}$, The $\phi(c_i)$ express by the comprehensive evaluation grade of evaluation science and technology information products that belongs to c_i 's reliability. c_i means the quantitative value of grade c_i , $c_{i+1} > c_i$, and then $\phi(c_i)$ is unascertained measure. Yet $[[c_1, c_5], \phi(c)]$ is an unascertained number. We call $E = \sum_{i=1}^5 c_i \phi(c_i)$ as unascertained measure expectation.

Unascertained measure expectation reflects the degree of science and technology information product social benefits. According to expected value we decide the social benefits of science and technology information products, the more expectation value, and the better social benefits.

Regarding the rank quantization value experts can ground the specific circumstances to different numerical, such as the quantitative value of c_5 is 90, the quantitative value of c_4 is 75, the quantitative value of c_3 is 60, the quantitative value of c_2 is 45, the quantitative value of c_1 is 30 and so on. We can figure out unascertained measure expectation E according to given quantitative value. When $90 \geq E \geq 81$, it has been thought that the comprehensive social benefits of the science and technology information products is very good. When $80 \geq E \geq 70$, it has been thought that the comprehensive social benefits of the science and technology information products is good. When $69 \geq E \geq 56$ it has been thought that the comprehensive social benefits of the science and technology information products is normal. When $55 \geq E \geq 45$, it has been thought that the comprehensive social benefits of the science and technology information products is worse. When $E < 45$, it has been thought that the comprehensive social benefits of the science and technology information products is bad.

4 Real Example of Explanation

“Super miniature camera technology” developed by one company in Shenzhen is a typical science and technology information product. It showed great economic and social benefit. This article will evaluate its social benefit.

Before the evaluation, it is necessary for us to evaluate some characteristics of the development and application of this high-tech information product.

(1) this technology is an electronic visual system which is made up by CMOS image sensor technology, computer image processing, automatic control and other technology, it has became an important corollary equipment that apply to the industry, business, public security, military, finance, trafficculture and education, medical treatment, private house monitoring and some other occasions. But in our country and other developing countries, the configuration of the system is still in the initial stage.

(2) The camera made by this technology is cheaper, smaller and lighter than the present common CCD camera.

(3) Synthesize camera developing prospect of domestic and overseas, developing a kind of CMOS super miniature camera will be able to lead to various electronic visual high technology products. Because this type of CMOS super miniature television camera has the advantages of small volume, low price, nice performance, energy saving, strong adaptability ,easy to use and so on, it will make TV technology in the application status of the industry and civil vision have epochal revolution.

(4) The main technology index and basic function is superior to CCD camera, it fills the gap of CMOS camera production in China.

The main technical and economic indexes:

The volume of this camera is one-fifth of the current micro CCD camera and one-twentieth of the ordinary CCD camera. Compare with the CCD camera of the same performance index the price is low 40 percent, and it saves the power.

The camera made by this technology is cheaper, smaller and lighter than the current CCD camera. According to the content above, we determine six evaluation factors:

f_1 : promote the progress of science and technology, f_2 : promote economic development, f_3 : strengthen social stability, f_4 : save social resources, f_5 : reduce the burden of consumption, f_6 : boost decision-making standard, f_7 : satisfy security needs, f_8 : active cultural life, f_9 : fill gaps of the market, f_{10} : improve nation's reputation. Invite 12 experts to determine each weight of index,

$w = (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9, w_{10}) = \{0.15, 0.15, 0.15, 0.1, 0.1, 0.06, 0.06, 0.07, 0.1, 0.06\}$, and evaluate each factor, getting the weights of evaluation factors and evaluation criteria table as following:

Table 1. Evaluation factor weight and evaluation scale table

evaluationf actors	(w)	evaluationscale				
		C ₅ (90)	C ₄ (75)	C ₃ (60)	C ₂ (45)	C ₁ (30)
f ₁	0.15	0.6	0.4	0	0	0
f ₂	0.15	0.3	0.5	0.2	0	0
f ₃	0.15	0.5	0.4	0.1	0	0
f ₄	0.1	0.4	0.4	0.2	0	0
f ₅	0.1	0.4	0.4	0.2	0	0
f ₆	0.06	0.3	0.3	0.4	0	0
f ₇	0.06	0.2	0.4	0.2	0.2	0
f ₈	0.07	0.2	0.3	0.4	0.1	0
f ₉	0.10	0.3	0.4	0.3	0	0
f ₁₀	0.06	0.2	0.3	0.4	0.1	0

Hence the measure matrix is:

$$R = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 \\ 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0.5 & 0.4 & 0.1 & 0 & 0 \\ 0.4 & 0.4 & 0.2 & 0 & 0 \\ 0.4 & 0.4 & 0.2 & 0 & 0 \\ 0.3 & 0.3 & 0.4 & 0 & 0 \\ 0.2 & 0.4 & 0.2 & 0.2 & 0 \\ 0.2 & 0.3 & 0.4 & 0.1 & 0 \\ 0.3 & 0.4 & 0.3 & 0 & 0 \\ 0.2 & 0.3 & 0.4 & 0.1 & 0 \end{pmatrix} \quad (2)$$

$$W = (0.15, 0.15, 0.15, 0.1, 0.1, 0.06, 0.06, 0.07, 0.1, 0.06) \quad (3)$$

$$S = WR = (0.376, 0.386, 0.203, 0.025, 0) = (\varphi(c_5), \varphi(c_4), \varphi(c_3), \varphi(c_2), \varphi(c_1)) \quad (4)$$

$$c = (c_1, c_2, c_3, c_4, c_5) = (30, 45, 60, 75, 90) \quad (5)$$

$$E = \sum_{i=1}^5 c_i \varphi(c_i) = 76.095, 80 \geq E \geq 70 \quad (6)$$

The above mentioned method using the unascertained measure and unascertained expectation model comprehensively evaluate the social benefits of "super miniature camera technology", we have know that the remarkable social benefit of this information product is good and considerable. Therefore, it is worth popularization and application.

References

1. Zhang, R.: Study on the post-evaluation of social economic benefits of logistics projects based on fuzzy AHP (analytic hierarchy process). *Technology and Method* (8), 62–63, 69 (2010)
2. Li, C., Li, H.: Multi-attributive group decision evaluation model under the condition of uncertain information. *Journal of XI'AN University of Architecture and Technology* (2), 111–115 (2010)
3. Gao, Y., Li, H., Wang, N.: Multi-objective fuzzy Trade-Off Optimization in Construction project under uncertainty. *Mathematics in Practice and Theory* (6), 152–159 (2010)
4. Li, J.: Fuzzy comprehensive evaluation of enterprise performance technology. *Information* (9), 811–812 (2010)
5. Sun, K., Jiang, M.: The fuzzy comprehensive evaluation method in the application of ChangJiu high speed railway project social benefits evaluation. *Journal of Engineering Management* (1), 49–52 (2012)
6. Zhao, G., Chen, H.: The unascertained measure model in the application of agricultural benefit evaluation. *An Hui Agricultural Science and Technology* (5), 2209–2210, 2215 (2010)
7. Dong, J., Pu, X.: Application of Unascertained number theory to communication effect appraisement. *Computer & Digital Engineering* (12), 159–161 (2008)
8. Wang, A.: The application of unascertained measurement model in building energy conservation project appraisal. *Mathematics in Practice and Theory* (7), 76–81 (2008)
9. Zhao, G., Liu, J.: Application of unascertained expectation in the decision making of agricultural investment. *Journal of Anhui Agricultural Science* (4), 1657–1670 (2010)
10. Wang, L., Wang, Q.: Base on the unascertained rational number in the reliability evaluation of transmission systems. *Shaanxi Electric Power* (07), 20–24 (2010)

Utility Analysis on Influence of Monetary Policy Adjustment on Jiangsu Province Export Trade in International Financial Crisis

Yue Ying¹, Ma Xing-hua², and Zhang Cheng-hu³

¹ Department of Finance, Tangshan Vocational & Technical College, Tangshan, Hebei, 063000, China

² College of Science, Hebei United University, Tangshan, Hebei, 063009, China

³ College of Languages, Hebei United University, Tangshan, Hebei, 063009, China
{yueyingts, zhangchenghu_jx}@126.com, mxh332@163.com

Abstract. Monetary policy is an important means to address the financial crisis. In the international financial crisis period, China has implemented a moderately loose monetary policy, which played a role to ease the financing difficulties of export enterprises. The paper uses event analysis which test and analyse the implemental effectiveness of deposit reserve ratio and the benchmark deposit and loan interest rate adjustment on the export trade of Jiangsu province.

Key words: Monetary Policy, Financial Crisis, Event Analysis.

1 Basic Principles of Event Analysis

Event analysis method is mainly used in financial field, which make researches on phenomenon of asset prices above quota yield which caused by accounting system adjustment, earnings announcements and trading rules changes. The key of this analysis method is the excess rate of return calculation and test. In the study of whether a certain event have impact or not on the financial market, its effectiveness is based on a basic assumption that the market is rational, event impact will react instantly in the asset price changes. The event analysis method for the study of international financial crisis policy factors in effectiveness validity based on this assumption: supposing the market is rational, the policy factors impacts will be reflected in the size and structure of export trade of Jiangsu province.

Event analysis method generally includes 7 steps:

- (1) Define and screen events.
- (2) Determine the event's publishing period, the window period and Estimation period.

The specific window settings are shown in Figure 1.

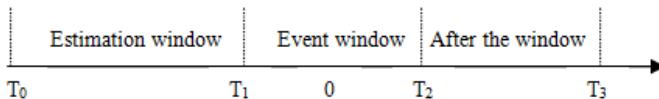


Fig. 1. Window settings

- (3) Sample selection.
- (4) Calculation of the normal rate of return, or if there is no occurrence of the event, how much return should be.

Generally we determine the normal rate in 4 ways, namely, the average rate of return method, market rate of return method, proxy portfolio return and risk adjusted return method. This paper uses the normal rate of return method to calculate event window in the normal income, according to the estimated window export trade average growth rate.

- (5) Calculation of abnormal return (AF), i.e. calculating events' real revenue minus normal returns.
- (6) Calculation of cumulative abnormal return (CAF).
- (7) Determine the statistical significance of AF and CAF.

2 Adjustment of the Main Monetary Policies in the International Financial Crisis Period

To counter the international financial crisis' influence on China's economy, the central government made "expanding domestic demand, growth, and development" as the keynote, based on expanding domestic demand, implement the moderately loose monetary policy, increasing finance to support strength, promoting steady and rapid economic development. The country's fiscal policy turned from tight to loose, and a number of related measures have been introduced. The publication of these measures put emphasis on financial support for economic development, widening financing channels and other aspects.

On November 10, 2008, the People's Bank of China put forward to strengthen the monitoring and analysis economic on financial operation, and deal with the work of financial services practically. At the same time, it stressed to solve the liquidity problems of the financial system, enhance the bank's credit in promoting economic development, and develop the bond market financing function further. Effective detection of financial operations ensures the smooth development of the financial institutions, and thus providing favorable financing environment for the enterprises.

On December 28, 2008, general office of the State Council issued on Several Opinions of the Current Financial Promotion of Economic Development, putting forward 30 opinions nine aspects, which involved "strengthen and improve credit services to meet the reasonable demand for capital; innovative financing means to broaden the financing channels for enterprises; improve foreign exchange management vigorously promote trade and investment facilitation", these observations provide strong financial support in order to ensure the growth of the export trade.

Along with the deepening of financial impact, the domestic enterprises, especially the real estate enterprises and small and medium enterprises' capital lack of liquidity caused financial crisis spread from the capital market to the real economy. The impact on Jiangsu Province export trade is enterprises financing difficulties. In order to establish confidence in the financial market, improve market liquidity, the People's Bank of China in the international financial crisis has lowered the benchmark of saving and loan interest rates, directly adjusting currency supply. From September to December, the People's Bank of China has lowered the deposit reserve ratio four times, as shown in table 1, table 2 and table 3.

Table 1. Situation of required reserve ratios reduced(a)

Published time	Effective date	Large financial institutions		
		Before	After	Range
September 15, 2008	September 25, 2008	17.50%	17.50%	0.00%
October 8, 2008	October 15, 2008	17.50%	17.00%	-0.50%
November 26, 2008	December 5, 2008	17.00%	16.00%	-1.00%
December 22, 2008	December 25, 2008	16.00%	15.50%	-0.50%

Table 2. Situation of required reserve ratios reduced(b)

Published time	Effective date	Small and medium-sized financial institutions		
		Before	After	Range
September 15, 2008	September 25, 2008	17.50%	16.50%	-1.00%
October 8, 2008	October 15, 2008	16.50%	16.00%	-0.50%
November 26, 2008	December 5, 2008	16.00%	14.00%	-2.00%
December 22, 2008	December 25, 2008	14.00%	13.50%	-0.50%

Table 3. Situation of benchmark interest rates for savings and loans cut

Timing data	benchmark interest rates for savings			benchmark interest rates for loans		
	Before	After	Range	Before	After	Range
September 16, 2008	4.14%	4.14%	0.00%	7.47%	7.20%	-0.27%
October 9, 2008	4.14%	3.87%	-0.27%	7.20%	6.93%	-0.27%
October 30, 2008	3.87%	3.60%	-0.27%	6.93%	6.66%	-0.27%
November 27, 2008	3.60%	2.52%	-1.08%	6.66%	5.58%	-1.08%
December 23, 2008	2.52%	2.25%	-0.27%	5.58%	5.31%	-0.27%

In adjusting the deposit reserve rate at the same time, the central bank lowered five times the benchmark deposit and loan interest rates, as shown in table 3.

The deposit reserve ratio and the benchmark interest rates for savings and loans cut manifests the government and financial regulators implement the moderately loose monetary policy flexibly and prudently, which provides good foundation for obtaining capital buffer for the export trade of Jiangsu province.

3 Influence of Monetary Policy Adjustment on Jiangsu Province' Export Trade

In theory, the deposit reserve ratio and the benchmark interest rates for savings and loans cut can increase the money supply, increase the liquidity of the banking system and make the export enterprises to broaden the financing channels. This section respectively examines the utility of monetary policy's adjustment on increasing export of main trader by using the event study method.

3.1 The Event Window and Prediction Window Selection

This section will regard monetary policy adjustment as the research object, and make export excess growth of foreign funded enterprises, private enterprises, state-owned enterprises, collective enterprises in Jiangsu Province after seasonal adjustment as measure index. Considering the intensive degree of incident released, policy lag and avoiding the peak of growth in October, 2008 and the most serious impact of international financial crisis in February 2009, we make April, 2009 to October, 2009, as the event window. There is acuteness and fluctuant in estimation window period, so trying to select as long as enough, for 36 months before the event window.

3.2 Empirical Model

The so-called abnormal returns are the actual value subtracting from the normal value (normal distribution) in the event window, the calculation formula is as follows:

$$AF_{i,t} = F_{i,t} - CF_i, \quad t = T_1, T_1 + 1, \dots, T_2 \quad (1)$$

Among them, i stands for event, and t is calculated by month, $AF_{i,t}$ expresses the abnormal turn of event i in the first month t . The event i corresponding to the seasonally adjusted trade main exports in the first month t are expressed by $F_{i,t}$. CF_i Stands that event i corresponds to the normal value of commodity exports in the first month t . Normal value is calculated by the average growth rate in estimated window. Then calculate the cumulative abnormal return value $CAF_{i,t}$ which is date t corresponding to the event i in event window. The formula is as follows:

$$t = T_1, T_1 + 1, \dots, T_2 \quad (2)$$

Set the number of enterprise species is N , for all enterprises, average cumulative abnormal change of the export volume in stage t of the event window is:

$$CAF_{i,t} = \sum_{T_1}^{T_2} AF_{i,t}$$

In order to verify the effects of main exports trader caused by monetary policy adjustment in the event window further, we take $\{\overline{CAF}_t\}$ as samples and analyzed by t test. And set $X_t = \overline{CAF}_t$, then, $\bar{X} = \frac{1}{T_2 - T_1 + 1} \sum_{t=T_1}^{T_2} X_t$

The original hypothesis (H_0): In the event window, monetary policy adjustment has no effect on exports of the main trader. That is $\bar{X} = 0$.

Alternative hypothesis (H_1): $\bar{X} \neq 0$

$$\text{Test statistics for } t(n-1) = \frac{\bar{X}}{S / \sqrt{n}}$$

$$\text{Where } S \text{ is the sample variance, } S = \sqrt{\frac{1}{n-1} \sum_{t=T_1}^{T_2} (X_t - \bar{X})^2}$$

N is the length of event window, that is $n = T_2 - T_1 + 1 = 7$.

3.3 The Results of Empirical Test

On the test of monetary policy adjustment on Jiangsu Province export trade main body of abnormal growth in international financial crisis, which can examine whether the support effect of monetary policy adjustment on the main foreign trade of Jiangsu province is remarkable or not in financial crisis. The calculation results as shown in Figure 2, the horizontal axis represents event window, and we can see that Jiangsu Province's main trade exports average cumulative abnormal growth is a positive trend in the whole event window period. But the results of test show that, $t_{(6)} = -0.6103$, at a

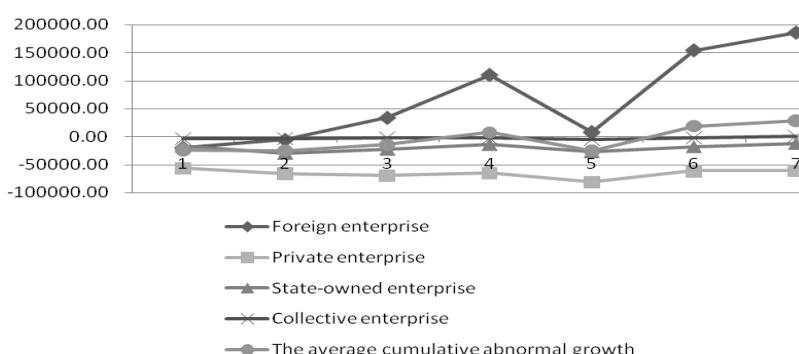


Fig. 1. Average cumulative excess growth of main trade enterprises exports in Jiangsu Province

75% confidence level, on the condition that the degree of freedom is 6, $|t| < 0.7161$, which is decide to accept a null hypothesis, namely, monetary policy adjustment's effects on the size of main export trader in Jiangsu province are not significant and effective.

4 Conclusion and Analysis

(1) During the international financial crisis, the stimulative effects of monetary policy adjustment on the size of main export trader in Jiangsu province are not obvious. This may be due to the function of monetary policy to the export trade is relatively limited. It can ease operating pressure of export enterprises in a short period of time, but not solve the difficulties which export trade is facing with fundamentally.

(2) Monetary policy has reflected certain support function in the export trade of Jiangsu province by the tremendous impact of the first half year, and the excess growth of different traders all reflects the upward trend. For foreign-funded enterprises, functions are relatively larger, but it still did not play a role in the private enterprises export, possibly due to dealing with private enterprise credit, the financial institutions would be more prudent.

By utility inspection of fiscal policy and monetary policy to the export trade of Jiangsu province during international financial crisis period, obviously, state governments at all levels made great efforts to support export trade and deal with financial crisis. Compared with the effectiveness of monetary policy, fiscal policy has provided more favorable measures. Among them, the adjustment of export tax rebate policy plays a leading role on export expansion, but it is not significant on statistics; It plays an effective role in the adjustment of main export commodity structure, which explain the government, improving the export tax rebate to help export trade to resist the financial crisis, at the same time, to take into account the export commodity structure optimization, paying more attention to the long-term development of export trade. While the deposit of reserve ratio and the cut of benchmark and loan interest rate have no effective improvement of main export financing environment, or for more indirect interpretation.

References

- Persaud, A.D.: Liquidity Black Holes: Understanding, Quantifying and Managing Financial Liquidity Risk. China Financial Publishing House, Beijing (2007)
- Ted Hong, C.H.: Handbook of Quantitative Finance and Risk Management. Springer, US (2010)
- Gunaward, P.: The Asia Currency Crisis and Australian Exports to East Asia. Economic Ananlysis and Policy 35, 73–90 (2008)
- Gnan, M.: The Impact of the Financial Crisis on the Real Economy in Austria—Analytical Challenges from Exceptional Factors. Monetary Policy and the Economy 4, 5–7 (2008)
- Hayford, M.D., Malliaris, A.G.: Asset Prices and the Financial Crisis of 2007–09: An erview Theories and Policies. Forum for Social Economics (2010)

6. Pressacco, F., Seravalli, G.: The Real Origins of the Global Crisis. *Transition Studies Review* 10 (2009)
7. Gros, D., Alcidi, C.: The impact of the financial crisis on the real economy. *Intereconomics*. 1(1), 45 (2010)
8. Eickmeier, S., Hofmann, B.: Monetary Policy, HousingBooms and Financial (Im) Balances. ECB Working Paper No. 1178 (2010)
9. Tempelman, J.H.: The financial crisis and the implementation of monetary policy. *Business Economics* 44, 216–219 (2009)
10. Dornbusch, R., Park, Y.C., Contagion, C.S.: Understanding How It Spreads. *World Bank Research Observer* 15, 177–197 (2000)

Author Index

- Abraha-Weldemariam, Daniel II-422
Ai, Chi II-709
An, Hongjing I-294
An, Hui I-44
An, Ran II-552
An, Yongli I-569

Bai, Jing II-590
Bao, Jutai II-604

Cai, Xi II-98, II-106
Cang, Dingbang II-334
Cao, Fukai II-21, II-249
Cao, Huirong I-280, II-414
Cao, Yuan-Yuan I-476
Chang, Jian II-590
Chang, Yan-shuo I-90
Chen, Danjiang II-611
Chen, Hang I-548
Chen, Jianli I-833
Chen, Jiaqin II-243
Chen, Junyu I-762
Chen, Liang-Liang I-659, I-795
Chen, Min-Rong I-216
Chen, Wenquan I-104
Chen, Wuping II-227
Chen, Xiai I-104
Chen, Xiang-yu I-273
Chen, Xueyou II-379
Chen, Yan-guang II-828
Chen, Ying I-694
Chen, Yingwei I-360
Chen, Yiwu I-225
Chen, Yong II-775
Chen, Yue II-195
Chen, Zhiyu I-427
Cheng, Jian II-15
Cheng, Meihua I-265, II-236
Chu, Jinxiu II-583, II-590
Cong, Shuai I-1
Cui, Chun-Sheng I-523
Cui, Delong II-146
Cui, Yanqiu I-183
Cui, Yanru I-200

Deng, Wei I-208
Ding, Jifeng I-183
Ding, Wenrui II-254
Dong, Bo I-686
Dong, Hong-bin I-716
Dong, Lu II-508
Dong, Ping I-161
Dong, Wenlei I-360
Dong, Xueping I-232, I-249
Dong, Zhangzhuo I-747
Dou, Wenhua I-538
Du, Qiao-qiao I-273
Du, Shujie I-16, I-484
Duan, Zhipeng II-685

Fan, Limei I-507
Fan, QiuHong II-92, II-309, II-678
Fang, Gang I-175
Fang, Yuejing II-791
Feng, Fuping II-709
Feng, Hai-Lin I-476
Feng, Jianguo I-288
Feng, Zhifang I-368, II-315, II-791
Fu, Fenglian I-336
Fu, Zhenlong II-67
Fu, Zhibin I-722

Gao, Aili I-762
Gao, Feng I-448
Gao, Ge I-273
Gao, Hongbo II-171
Gao, Jiansheng II-21
Gao, Luyang II-27
Gao, Min II-643
Gao, Rujia I-37, I-427
Gao, Shanqing II-658
Gao, Wei II-737
Gao, Wenhui I-604
Gao, Xuefei I-569
Gao, Yan I-329, I-421, I-624
Ge, Bo I-138
Ge, Shigang II-334
Gong, ShengWen I-610
Gong, Wenjing II-604
Gu, Jiantao II-34, II-41, II-47

- Gu, Lichuan II-387
 Gu, Zhao II-34, II-41, II-47
 Guan, Shuming II-451
 Guan, Zhitao I-192
 Gui, Wenyong II-341
 Gui, Zhanji II-544
 Guo, Fangfang I-161
 Guo, Hong I-783
 Guo, Jing I-192
 Guo, Lizheng II-119
 Guo, Pengfei I-433
 Guo, Ping II-396
 Guo, Wei I-783
 Guo, Wensheng I-58
 Guo, Ying I-515
 Guo, Yuanbo II-651
 Guo, Zhe I-469
 Guo, Zhirong II-737
 Guo, Zu-hua II-480
- Han, Bing I-387, II-496, II-502
 Han, Bo I-169
 Han, Gang I-538
 Han, Guang II-98, II-106
 Han, Jian II-575
 Han, Shuying II-583, II-590
 Han, Ting II-583, II-590
 Hao, Xiaohong I-582
 Hao, Zhi-feng II-693
 He, Jun II-701
 He, Lijun I-352
 He, Lingna I-225
 He, Songhua I-125
 He, Xiaohu II-422
 He, Xiaoting I-747
 He, Yigang I-232, I-249
 He, Zongyao II-119
 Hou, Lifeng I-44
 Hu, Guyu I-82
 Hu, Jianbin I-576
 Hu, Lin II-112
 Hu, Ling I-175
 Hu, Ming I-427
 Hu, Xinxin II-219
 Hu, Xulei II-766
 Hua, Qing-yi I-90
 Hua, Rong II-611
 Hua, Xiaojun II-737
 Huang, Feng II-67
 Huang, Han II-693
- Huang, Jinfeng I-152
 Huang, Zhiqin I-653
 Huang, Zhongsheng II-315, II-791
- Ji, Junzhong II-404
 Ji, Nan II-288
 Jia, Rui I-708
 Jiang, Changyun II-119
 Jiang, Jiulei II-804
 Jiang, Junna II-502
 Jiang, Tongsong II-356, II-371
 Jiang, Yan II-583, II-590
 Jiang, Zetao I-200
 Jiang, Zhao-Lin II-364
 Jiang, Ziwu II-356
 Jiao, Cui-Ling I-25, I-31, I-756
 Jiao, Lang II-404
 Jiao, Ling I-484
 Jing, Haibin II-300
 Jing, Lin II-219
- Ke, Dongliang I-9
 Ke, Shuang I-104
- Lai, Jinhui II-146
 Lei, Mei-Shen I-597
 Lei, Tianyu II-528
 Li, Baoliang I-538
 Li, Cheng II-567
 Li, Chenxia I-433
 Li, Dan I-672
 Li, Dongmei I-421
 Li, Fang I-74
 Li, Gen I-463
 Li, Guihua I-288
 Li, Haixia I-9
 Li, Hang I-37, I-638
 Li, Heling II-262
 Li, Hongguang II-254
 Li, Huabo I-82
 Li, Huifen I-820
 Li, Jianbo I-66
 Li, Jianjun II-227
 Li, Jianxin I-747
 Li, Jiaqi I-582
 Li, Jie II-635
 Li, Ji-Lin II-152
 Li, Jing II-701
 Li, Jinzhao I-722
 Li, Ji-Rui I-51, I-441

- Li, Juan II-364
 Li, Kun I-747
 Li, Li II-348
 Li, Lihong I-329
 Li, Linfan I-624
 Li, Mei II-724
 Li, Mingduo II-429, II-766
 Li, Ming-zhong I-74
 Li, Peng II-92, II-309
 Li, Qingbao II-171
 Li, Qingshui I-225
 Li, Shan I-97
 Li, Shaowen II-387
 Li, Shiyong I-456
 Li, Suhong II-280
 Li, Wanlong I-37, I-638
 Li, Wei I-192, I-273, II-834
 Li, Wei-hua II-53, II-438
 Li, Wensheng I-368, II-791
 Li, Xia I-216
 Li, Xiang I-523
 Li, Yao I-175
 Li, Ying I-399, II-85
 Li, Yuanxiang II-466
 Li, Yue I-110, I-117
 Li, Zhengwen I-169
 Li, Zhiyong I-294, II-685
 Lian, Lijun II-61
 Liang, Song II-146
 Liao, Zhijiang II-643
 Lin, Deyu I-531
 Lin, Rongju I-208
 Liu, Baoxiang II-309
 Liu, Caiming I-576, II-488
 Liu, Chunnnian II-404
 Liu, Dan II-480
 Liu, Fangfang I-646
 Liu, Feng I-795
 Liu, Fenghua I-659
 Liu, Fengling I-653
 Liu, Fuchun I-392
 Liu, Haisheng II-341
 Liu, Hongli II-685
 Liu, Huiqin I-630, II-730
 Liu, Huixin I-448
 Liu, Junqiang I-427
 Liu, Lanfang I-484
 Liu, Lijuan II-665
 Liu, Renjin II-1
 Liu, Tao II-528
 Liu, Wensheng II-759
 Liu, Xin II-322
 Liu, Yajing II-724
 Liu, Yanli I-382, II-272
 Liu, Yixue I-783
 Liu, Yong II-171
 Liu, Yongchang II-752
 Liu, Yonghong I-208
 Liu, Yongkui I-352
 Liu, Yongqiong II-552
 Liu, Yutao I-515
 Liu, Zhenpeng I-1
 Liu, Zhijun II-404
 Liu, Zhiyou I-448
 Liu, Zhujin II-396
 Lu, Aiying II-254
 Lu, Jia I-538
 Lu, Ming I-408
 Lu, Shan I-242
 Lu, Xiaoqing II-451
 Lu, Zhaoquan I-232, I-249
 Luo, Dan I-789
 Luo, Jianping I-216
 Luo, Shihua I-392
 Lv, Shanguang I-820
 Lv, Yun-hong I-678
 Ma, Dong II-67
 Ma, Hongjiang II-76
 Ma, Huixia II-604
 Ma, Jun II-651, II-752
 Ma, Junhong I-387, II-288, II-451, II-496
 Ma, Liang I-795
 Ma, Qiongyu I-368
 Ma, Shi-xia II-480
 Ma, Xiaodan II-133
 Ma, Xing-hua I-840
 Mao, Fang II-76
 Meng, Mingchuan I-590, I-827
 Meng, Xiaojing I-531
 Mi, Xue-yu I-686
 Miao, Jianying II-164
 Miao, Wenjing II-328
 Miao, Zuohua II-775
 Min, Chen Rui II-812
 Ni, Yunlu I-548
 Nie, Huaiyong II-445, II-508
 Ning, Jing II-219

- Niu, Chuncheng I-492
 Niu, Jian-Zhao II-152
 Ouyang, Yuanling I-492
 Pan, Wei I-469
 Pang, Nan II-158
 Pei, Hongwen II-628
 Peng, Chengbao I-708
 Peng, Lingxi I-576, II-488
 Peng, Tao I-665
 Peng, Xinyi I-638
 Pu, Changjiu I-500
 Qian, Li I-562
 Qian, Lijun II-821
 Qu, Bochao II-348
 Qu, Dong I-313
 Qu, Mingyang II-138
 Rao, Hong I-97
 Ren, Minglun II-211
 Ruan, Guangce II-783
 Ruan, Yu-hua I-273
 Sang, Nan I-58
 Shan, Zhao II-812
 Shang, Kaiyu I-1
 Shao, Shengchun I-702
 Shao, Yanling I-138
 Shen, Dingcai II-466
 Shen, Hong I-110, I-117
 Shen, Nuo II-364
 Shi, Li I-716
 Shi, Linan II-348
 Shi, Qin II-821
 Shi, Shengli I-722
 Shi, Yaqing I-9
 Shu, Yunxing I-138, I-773
 Song, Liufu I-427
 Song, Weidong I-646
 Song, Wentao II-187
 Song, Yibing II-15
 Su, Hongsheng I-729
 Sun, Hongkai II-300
 Sun, Jian I-110
 Sun, Jiangsheng II-671
 Sun, Wei I-456
 Sun, Xu I-810
 Tan, Lu I-373
 Tan, Yili I-382, II-322
 Tang, Jundong II-27
 Tang, Weiwen I-576
 Tian, Guanhua I-192
 Tian, Jia-Li II-152, II-567, II-575, II-596
 Tian, Jingrui I-630, II-730
 Tian, Xiaojie I-208
 Tong, Wei I-321
 Wang, Ai-Jing II-693
 Wang, Aimin II-635
 Wang, Aiming I-659
 Wang, Cilin II-488
 Wang, Degao I-352
 Wang, Guochun I-638
 Wang, Hao I-66
 Wang, Hong I-258, I-694, II-521
 Wang, Hong-jiang I-802
 Wang, Hongli I-630, II-730
 Wang, Hong-rui I-399
 Wang, Jiangjing II-804
 Wang, Jian-Ping I-25, I-31, I-756
 Wang, Jiawei II-187
 Wang, Jiaxing II-254
 Wang, Jierui I-618
 Wang, Jin I-133
 Wang, Jinfang I-117
 Wang, Jingbo I-562
 Wang, Jing-sheng I-399
 Wang, Jingtao I-562
 Wang, Jinkuan II-98, II-106
 Wang, Jinshuang II-457
 Wang, Jun-chao I-665
 Wang, Junhao II-119
 Wang, Junhui I-538
 Wang, Kaihua II-544
 Wang, Lijin II-219
 Wang, Ling I-104
 Wang, Peipei II-429
 Wang, Qi I-336
 Wang, Qiang I-200, II-658
 Wang, Qiangqing I-618
 Wang, Qing II-293
 Wang, Shaoguang II-671
 Wang, Shouping I-146
 Wang, Shuan-zhong I-463
 Wang, Shujun II-496
 Wang, Tao II-293
 Wang, Xian I-66
 Wang, Xiaolei II-356, II-371
 Wang, Xiaoyan II-138, II-561

- Wang, Xinchun II-502
 Wang, Xinxin II-745
 Wang, Xiu-lan II-480
 Wang, Xiuxia I-820
 Wang, Ya II-604
 Wang, Yan I-329
 Wang, Ying-yan I-74
 Wang, Yinlong II-709
 Wang, Yinqiu II-622
 Wang, Yongjun II-67
 Wang, Yourong I-382
 Wang, Zhenbing I-146
 Wang, Zhijun I-360
 Wang, Zhilu II-583
 Wang, Zhuxiao I-192
 Wang, Zongjiang II-203
 Wei, Bo II-466
 Wei, Cheng-zhi I-665
 Wei, Guang-Hui I-659, I-795
 Wei, Jing II-341
 Wei, Li-xin I-399
 Wei, Qun II-309, II-678
 Wei, Wei II-536
 Wei, Xiao-peng I-802
 Wei, Yongxia II-138, II-561
 Wen, Deliang II-133
 Wen, Jiechang I-336
 Wu, Bo I-484
 Wu, Junjun I-590, I-827
 Wu, Kehe I-192
 Wu, Lihua I-392
 Wu, Qingyong I-9
 Wu, Tao II-834
 Wu, Xiaoyun II-658
 Wu, XueLi I-44
 Wu, Yan II-34
 Wu, Yun I-313, I-321
 Wu, Yun-zhang II-717
 Xia, Bin I-827
 Xia, Jun-tie I-802
 Xia, Le I-288
 Xia, Zhi-qiu I-802
 Xiao, Benluo II-41, II-47
 Xiao, Mingbo I-556
 Xiao, Xin II-280
 Xing, Huaming I-368, II-315, II-791
 Xing, Lixin II-658
 Xiong, Jiang I-175
 Xiong, Yan II-1
 Xiong, Ying II-262
 Xu, Bo I-433
 Xu, Chunxiang II-422
 Xu, Dengyuan II-521
 Xu, Hong I-104
 Xu, Jing I-702, II-514
 Xu, Shicheng I-708
 Xu, Shouzhi I-433
 Xu, Shuang I-183
 Xu, Xiaosi I-694
 Xu, Xing II-466
 Xu, Xinwen I-345
 Xu, Yan-mei II-596
 Xu, Yingzhuo I-738
 Xu, Yulong I-208
 Yan, Pengcheng II-671
 Yan, Shaohong I-306
 Yan, Su II-671
 Yang, Bin II-262
 Yang, Guangsong I-556
 Yang, Guan-Nan I-523
 Yang, Guowu I-242
 Yang, Hai-bo I-414
 Yang, Hongmei II-322
 Yang, Hongyu II-300
 Yang, Jin II-488
 Yang, Jing II-474
 Yang, Jiongzhao II-92
 Yang, Panpan II-211
 Yang, Pingping II-429, II-766
 Yang, Shuangqi I-702, II-514
 Yang, Xiang I-469
 Yang, Yong-Jian I-523
 Yang, Yun I-82
 Yang, Zhihui I-161
 Yao, Jie II-717
 Ye, Jin-tao I-665
 Ye, Suping II-474
 Ye, Yinzhong II-611
 Ye, Yong II-387
 Yin, Jindan II-709
 Yin, Lisheng I-232, I-249
 Yin, Zhi-yuan I-665
 Yong, Xiao II-812
 Yu, Bin II-227
 Yu, Fei II-466
 Yu, Hong II-583, II-590
 Yu, Kui II-514
 Yu, Le II-488

- Yu, Ning I-414
 Yu, Wu I-66
 Yu, Xiuping II-300
 Yu, Yaping II-85, II-272
 Yu, Yong I-345
 Yuan, Lina I-604
 Yuan, Shifang I-300
 Yuan, Zhijian II-457
 Yue, Lina II-445, II-508
 Yue, Mengtian I-280, II-414
 Yue, Ying I-840
 Yun, Jian I-352
 Yun, Shiwei I-773

 Zeng, Jinquan I-576
 Zeng, Rui I-74
 Zeng, Shouzhen I-265, II-236, II-243
 Zeng, Xiangyang II-775
 Zeng, Xianquan II-628
 Zhan, Jinyu I-242
 Zhang, Aidong II-404
 Zhang, Bin I-1
 Zhang, Cheng-hu I-840
 Zhang, Chuanlin I-762
 Zhang, Dong II-701
 Zhang, Haibo I-646
 Zhang, Haicheng II-9
 Zhang, Han I-373
 Zhang, Hao II-92, II-678
 Zhang, Heng II-552
 Zhang, Jian II-85, II-272
 Zhang, Jinan II-709
 Zhang, Jintang I-288
 Zhang, Jiuyun II-528
 Zhang, Lianwu II-671
 Zhang, Lin I-686
 Zhang, Lingmin II-280
 Zhang, Na II-119
 Zhang, Peng I-686
 Zhang, Qiansheng I-392
 Zhang, Rui I-208
 Zhang, Ruicheng II-187, II-429, II-766
 Zhang, Ruihong II-635
 Zhang, Shaobai II-195
 Zhang, Shouhua 1
 Zhang, Shuqin I-556
 Zhang, Sujuan II-752, II-834
 Zhang, Su-Qing I-51, I-441
 Zhang, Tao II-457
 Zhang, Tian I-597

 Zhang, Wenxiang II-544
 Zhang, Wenyi II-451
 Zhang, Yanli II-752, II-834
 Zhang, Yan-yan I-414
 Zhang, Ye-qing II-575
 Zhang, Yongli II-179
 Zhang, Yongsheng I-306
 Zhang, Yuanyi I-469
 Zhang, Yun-Peng I-672
 Zhang, Zhaozhong II-371
 Zhao, Guoyong I-294
 Zhao, Hui I-97
 Zhao, Huijuan II-280
 Zhao, Hui-qin II-596
 Zhao, Huiqin II-567
 Zhao, Jian-hua II-53, II-438
 Zhao, Jianlin II-745
 Zhao, Min II-457
 Zhao, Shuguang II-119
 Zhao, Ting-ting I-802
 Zhao, Wenjing II-348
 Zhao, Xin II-583, II-590
 Zhao, Yao II-249
 Zhao, Ying I-672
 Zhao, Yuyu II-561
 Zheng, Jun I-694, II-21, II-27
 Zheng, Shanhong I-37
 Zheng, Xia II-635
 Zheng, Xiao I-9
 Zheng, Xiuyun I-653
 Zheng, Zheng I-110, II-249
 Zhou, Ji II-552
 Zhou, Lihui I-258, I-306, I-421, I-624
 Zhou, Rui I-58
 Zhou, Wuzhong I-789
 Zhou, Xiancun II-1
 Zhou, Xiangbing II-76
 Zhou, Zhaoming II-701
 Zhou, Zhuan I-484
 Zhu, Kaiyu I-569
 Zhu, Lin II-396
 Zhu, Xiaohua II-127
 Zhu, Yanwei II-797
 Zhu, Yong-hong II-717
 Zhu, Yu II-171
 Zhu, Yu-tian I-463
 Zhuo, Xu II-812
 Zou, Kuancheng I-492
 Zou, Yi I-492