

Milan Tuba
Shyam Akashe
Amit Joshi *Editors*

Information and Communication Technology for Sustainable Development

Proceedings of ICT4SD 2018

Advances in Intelligent Systems and Computing

Volume 933

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences,
Warsaw, Poland

Advisory Editors

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing,
Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagras, School of Computer Science & Electronic Engineering,
University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University,
Gyor, Hungary

Vladik Kreinovich, Department of Computer Science, University of Texas
at El Paso, El Paso, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao
Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology,
University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute
of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro,
Rio de Janeiro, Brazil

Ngoc Thanh Nguyen, Faculty of Computer Science and Management,
Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering,
The Chinese University of Hong Kong, Shatin, Hong Kong

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at <http://www.springer.com/series/11156>

Milan Tuba · Shyam Akashe · Amit Joshi
Editors

Information and Communication Technology for Sustainable Development

Proceedings of ICT4SD 2018



Springer

Editors

Milan Tuba
Singidunum University
Belgrade, Serbia

Amit Joshi
Sabar Institute of Technology
Gujarat Technological University
Ahmedabad, Gujarat, India

Shyam Akashe
Department of Electronics
and Communication Engineering
ITM University
Gwalior, Madhya Pradesh, India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-981-13-7165-3

ISBN 978-981-13-7166-0 (eBook)

<https://doi.org/10.1007/978-981-13-7166-0>

Library of Congress Control Number: 2019934355

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721,
Singapore

Organization

Organizing Chairs

Mr. Bharat Patel, Vice Chairman, ASSOCHAM Western Region
Mr. Peter Kent, Chief Executive Officer (CEO), UKEI, UK

Organizing Co-Chair

Mr. Nakul Sharedalal, Advisor, G R Foundation, India

Organizing Secretary

Amit Joshi, SITG, Ahmedabad, and Global Knowledge Research Foundation

Members

Dr. Vijay Singh Rathore, Department of CSE, JECRC, Jaipur, India
Mr. Aman Barot, G R Foundation, India
Dr. Mahipal Singh Deora, BN University, Udaipur, India
Dr. Nisheeth Joshi, Banasthali University, Rajasthan, India
Mr. Nilesh Vaghela, Electromech, Ahmedabad, India
Mr. Vinod Thummari, SITG, Ahmedabad, Gujarat, India
Dr. Chirag Thaker, L D College of Engineering, Gujarat, India
Prof. S. Rama Rao, Goa University, Goa, India
Dr. Nitika Vats Doohan, Indore, India

Dr. Parikshit Mahalle, Sinhgad Group, Pune, India
Dr. Priyanks Sharma, RSU, Ahmedabad
Dr. Nitika Vats Doohan, Indore
Dr. Mukesh Sharma, SFSU, Jaipur
Dr. Manuj Joshi, SGI, Udaipur, India
Dr. Bharat Singh Deora, JRN RV University, Udaipur, India
Mr. Ricky, P. Tunes, Goa University, India
Prof. L. C. Bishnoi, GPC, Kota, India
Dr. Vikrant Bhatela, Lucknow, India
Dr. Satyen Parikh, Dean, Ganpat University, Ahmedabad, India
Dr. Puspendra Singh, JKLU, Jaipur, India
Dr. Aditya Patel, Ahmedabad University, Gujarat, India
Mr. Ajay Choudhary, IIT Roorkee, India

Technical Program Committee

Prof. Milan Tuba, Singidunum University, Serbia
Dr. Durgesh Kumar Mishra, Chairman, Div IV, CSI

Technical Program Committee Co-Chairs

Dr. Mahesh Bundele, Dean Research, Poornima University, Jaipur
Dr. Nilanjay Dey, Techno India College of Engineering, Kolkata, India

Advisory Committee

Shri. Nitin Kunkolienker, President, MAIT
Prof. S. Rama Rao, Goa University, Goa
Mr. P. N. Jain, Add. Sec., R&D, Government of Gujarat, India
Prof. J. Andrew Clark, Department of Computer Science, University of York, UK
Mr. Vivek Ogra, President, GESIA
Prof. Mustafizur Rahman, Endeavor Research Fellow, Australia
Dr. Kalpdrum Passi, Laurentian University, ON, Canada
Mr. Chandrashekhar Sahasrabudhe, ACM India
Dr. Pawan Lingras, Saint Mary's University, Canada
Dr. S. C. Satapathy, Professor in CSE, AP, India
Mr. Niko Phillips, Active Learning, UK, India
Dr. Bhagyesh Soneji, Chairperson, ASSOCHAM Western Region

Dr. Dharm Singh, NUST, Windhoek, Namibia
Dr. Vinay Chandna, Principal, JECRC, Jaipur, India
Prof. H. R. Vishwakarma, VIT, Vellore, India

Members

Dr. Amit Kaul, UK
Prof. Babita Gupta, College of Business, California State University, California, USA
Prof. Ting-Peng Liang, National Chengchi University, Taipei, Taiwan
Prof. Anand Paul, The School of Computer Science and Engineering, South Korea
Prof. Rachid Saadane, Casablanca, Morocco
Prof. Brent Waters, University of Texas, Austin, Texas, USA
Prof. Philip Yang, PricewaterhouseCoopers, Beijing, China
Prof. H. R. Vishwakarma, VIT, Vellore, India
Prof. Martin Everett, University of Manchester, England
Dr. Rajeev Vaghmare, Principal, SITG, Ahmedabad, Gujarat, India
Prof. XiuYing Tian, Instrument Laboratory, Yangtze Delta Region Institute of Tsinghua University, Jiaxing, China
Prof. Gengshen Zhong, Jinan, Shandong, China
Prof. Mustafizur Rahman, Endeavor Research Fellow, Australia
Prof. Ernest Chulantha Kulasekere, University of Moratuwa, Sri Lanka
Prof. Subhadip Basu, Visiting Scientist, The University of Iowa, Iowa City, USA
Dr. Ashish Rastogi, Higher College of Technology, Muscat, Oman
Prof. Ahmad Al-Khasawneh, The Hashemite University, Jordan
Dr. Basant Tiwari, Professor, Ethiopia University, Ethiopia
Prof. Jean Michel Bruel, Departement Informatique IUT de Blagnac, Blagnac, France
Dr. Ramesh Thakur, DAVV, Indore, India
Prof. Shuiqing Huang, Department of Information Management, Nanjing Agricultural University, Nanjing, China
Prof. Sami Mnasri, IRIT Laboratory Toulouse, France
Dr. Krishnamachar Prasad, Department of Electrical and Electronic Engineering, Auckland, New Zealand
Prof. Foufou Sebti, Dijon Cedex, France
Dr. Haibo Tian, School of Information Science and Technology, Guangzhou, Guangdong
Mr. Vinod Thummar, SITG, Ahmedabad, Gujarat, India
Prof. Sunarto Kampus, UNY, Yogyakarta, Indonesia
Prof. Feng Jiang, Harbin Institute of Technology, China
Prof. Feng Tian, Virginia Polytechnic Institute and State University, USA
Dr. Savita Gandhi, GU, Ahmedabad, India

- Prof. Abdul Rajak A. R., Department of Electronics and Communication Engineering, Birla Institute of Technology and Sciences, Abu Dhabi
Prof. Hoang Pham, Rutgers University, Piscataway, NJ, USA
Prof. Shashidhar Ram Joshi, Institute of Engineering, Pulchowk, Nepal
Prof. Abrar A. Qureshi, Ph.D., University of Virginia, USA
Dr. Aynur Unal, Stanford University, USA
Dr. Manju Mandot, JRN RV University, Udaipur, India
Dr. Rajveer Shekawat, Manipal University, India
Asst. Prof. Ngai-Man Cheung, Singapore University of Technology and Design, Singapore
Prof. J. Andrew Clark, Department of Computer Science, University of York, UK
Prof. Yun-Bae Kim, Sungkyunkwan University, South Korea
Prof. Lorne Olfman, Claremont, California, USA
Prof. Louis M. Rose, Department of Computer Science, University of York
Nedia Smairi, CNAM Laboratory, France
Dr. Vishal Gaur, Government Engineering College, Bikaner, India
Dr. Aditya Patel, Ahmedabad University, Gujarat, India
Mr. Ajay Choudhary, IIT Roorkee, India
Prof. Prasun Sinha, The Ohio State University, Columbus, OH, USA
Dr. Devesh Shrivastava, Manipal University, Jaipur, India
Prof. Xiaoyi Yu, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China
Prof. R. S. Rao, New Delhi, India

Preface

Third International Conference on ICT for Sustainable Development (ICT4SD 2018) targeted theory, development, applications, experiences, and evaluation of interaction sciences with fellow students, researchers, and practitioners.

Conference may concern any topic within the conference scope. Workshops may be related to any topics within the conference scope. The conference is devoted to increase the understanding role of technology issues, how engineering has day by day evolved to prepare human friendly technology. The conference provided a platform for bringing forth significant research and literature across the field of ICT for sustainable development and an overview of the technologies awaiting unveiling. This interaction was the focal point for leading experts to share their insights, provide guidance, and address participant's questions and concerns.

The conference was held during August 30–31, 2018, at Hotel Vivanta by Taj, Panaji, Goa, India, organized by Global Knowledge Research Foundation, state chamber partner—Goa Chamber of Commerce & Industry, and incubation partner—Centre for Incubation and Business Acceleration, and supported by The Institution of Engineers (India)—CEDB, IE(I); Department of Electronics and IT, Ministry of Communications & Information Technology—Government of India; Asian-African Chamber of Commerce & Industry; Unified Brainz; and World Peace & Diplomacy.

Research submissions in various advanced technology areas were received, and after a rigorous peer review process with the help of program committee members and 56 external reviewers for 400 papers from 8 different countries including Algeria, USA, United Arab Emirates, Serbia, Qatar, Mauritius, Egypt, Saudi Arabia, Ethiopia, and Oman, 82 were accepted with an acceptance ratio of 0.13.

Technology is the driving force of progress in this era of globalization. Information and communication technology (ICT) has become a functional requirement for the socioeconomic growth and sustainable development of any country. The influence of information and communication technology (ICT) in shaping the process of globalization, particularly in productivity, commercial, and financial spheres, is widely recognized. The ICT sector is undergoing a revolution that has momentous implications for the current and future social and economic

situation of all the countries in the world. ICT plays a pivotal role in empowering people for self-efficacy and in understanding how it can facilitate this mission to reach out to the grassroots level. Finally, it is concluded that ICT is a significant contributor to the success of the ongoing initiative of Startup India.

In order to recognize and reward the extraordinary performance and achievements by ICT and allied sectors and promote universities, researchers, and students through their research work adapting new scientific technologies and innovations, the two-day conference had presentations from the researchers, scientists, academia, and students on the research work carried out by them in different sectors.

ICT4SD & Start up Summit was a flagship event of G R Foundation. This was the third edition. The earlier two series were a grand success with participation from various universities, academia, scientists, scholars, researchers, students, industries, stakeholders, and R&D institutions from within and outside the country. The researchers presented their research papers through presentations at the conference. All the earlier conferences witnessed the presence of more than 250 delegates and eye-catching presence from government institutions and ministries. The third edition of ICT4SD & Start up Summit along with Startup & ICT Awards focused on the new innovations using scientific technologies in the ICT sector. The conference-cum-awards marked the presence of national and international stakeholders, universities, R&D institutions, academics, students from Goa University, and BITS Pilani, Goa Chapter, representatives from union and state governments, investors, policymakers, industry leaders, trade bodies and scientific consultants, and ICT experts to share their knowledge in this area.

Belgrade, Serbia

Gwalior, India

Ahmedabad, India

Milan Tuba

Shyam Akashe

Amit Joshi

Contents

Analysis of Performance Measures of Computer Systems with Priority and Maximum Operation Time	1
Indeewar Kumar, Ashish Kumar and Monika Saini	
Novel Concept of Spelling Correction for Semantic Tourism Search Interface	13
Shilpa Laddha and Pradip M. Jawandhiya	
Cost-Effective NSX Inventory State Management in Cross-Cloud	23
Saloni Garg and Ravindra Kumar	
A Tri-band Staircase Antenna for RF Energy Harvesting	33
Aysha Munir Sheikh and Rajeev Mathur	
Aspect-Based Sentiment Analysis Using Deep Learning Convolutional Neural Network	43
Ravindra Kumar and Saloni Garg	
BrownBoost Classifier-Based Bloom Hash Data Storage for Healthcare Big Data Analytics	53
S. Arun Kumar and M. Venkatesulu	
AmbuPod a Family-Owned Social Enterprise for Inclusive Quality Health care in India	71
Pradnya Vishwas Chitrapu, Pravin Kumar Bhoyar and (Brig) Rajiv Divekar	
Quality Assurance Practices and Techniques Used by QA Professional in Continuous Delivery	83
Anish Cherian, Raju Ramakrishna Gondkar and S. Suresh Babu	
Sensoponics: IoT-Enabled Automated Smart Irrigation and Soil Composition Monitoring System	93
Prasun Guchhait, Pranav Sehgal and Vidyadhar J. Aski	

A Novel Trust Scheme in Semantic Web	103
Suresh Kumar and Neera Chaudhary	
Random-opposition-based Learning for Computational Intelligence	111
Divya Bairathi and Dinesh Gopalani	
Secure Key Agreement Model for Group Data Sharing and Achieving Data Deduplication in Cloud Computing	121
Manav Ashok Thakur, Santosh Bari, Rutuja Deshmukh and Smita Auty	
Automated Road Surface Condition Monitoring System Using Machine Vision Technology	129
K. Sujatha, A. Ganeshan, V. Karthikeyan, P. Sai Krishna, Shaik Shafiya, N. P. G. Bhavani and V. Srividhya	
Evolution of Malware and Its Detection Techniques	139
Sanjay K. Sahay, Ashu Sharma and Hemant Rathore	
Preventing Forgeries by Securing Healthcare Data Using Blockchain Technology	151
V. Vetriselvi, Sridharan Pragatheeswaran, Varatharajan Thirunavukkarasu and Amaithi Rajan Arun	
A Survey of Deep Learning Techniques for Medical Diagnosis	161
Abdul Mueed Hafiz and Ghulam Mohiuddin Bhat	
Integrated Use of AHP and GIS Techniques for Selection of Artificial Water Recharge Sites	171
Shaikh Husen, Santosh Khamitkar, Parag Bhalchandra, Preetam Tamsekar, Govind Kulkarni and Kailas Hambarde	
Social Media Data Mining Techniques: A Survey	183
Dimple Tiwari and Manoj Kumar	
Automotive Vehicle-to-Everything (V2X) Communication Using IoT	195
P. Janani, Siddhant Verma, S. Natarajan and Aditya Kumar Sinha	
Summarization of Graph Using Question Answer Approach	205
Aditi P. Deshpande and C. N. Mahender	
Analytical Assessment of Nature-Inspired Metaheuristic Algorithms to Elucidate Assembly Line Task Scheduling Problem	217
Shantanu Anandrao Lohi, Ajay A. Jaiswal, Harish V. Gorewar and Snehal A. Lohi	
#BiggBoss—Long-Run Event Detection and Sentiment Mining in Twitter	229
R. Geetha, P. Suthanthira Devi and S. Karthika	

Land Use Land Cover Change Detection Through GIS and Unsupervised Learning Technique	239
Govind Kulkarni, Aniket Muley, Nilesh Deshmukh and Parag Bhalchandra	
Innovating Digital Consumer Services in Wealth Management Ecosystem an Anthropomorphic Approach	249
Aditya Kulkarni and Swapnil Dambe	
Security Challenges and Solutions of IoT Ecosystem	259
Anshul Jain and Tanya Singh	
A Wireless Sensor System to Monitor Banana Growth Based on the Temperature	271
B. Geethanjali and B. L. Muralidhara	
User Review Classification and Star Rating Prediction by Sentimental Analysis and Machine Learning Classifiers	279
Aagam Shah, Komal Kothari, Umang Thakkar and Satvik Khara	
Veracity Analysis and Prediction in Social Big Data	289
P. Suthanthira Devi, S. Karthika, P. Venugopal and R. Geetha	
Efficient Implementation of Carry-Skip Adder Using CSMT Adder and PPA	299
S. Sridevi Sathya Priya, Benita Poulraju, B. Benita and Dharani Karanatakam	
Multispectral Satellite Image Classification Based on Bare Bone Fireworks Algorithm	305
Eva Tuba, Raka Jovanovic and Milan Tuba	
A Survey of Directory and Database Protocols for Data Extraction	315
Manav Ashok Thakur, Santosh G. Bari, Rutuja Deshmukh and Smita Auty	
Recent Trends in ICT-Enabled Renewable Energy Systems.....	327
Shamkumar Chavan and Mahesh Chavan	
Assessing the Smart Readiness of Local Councils in Mauritius	333
B. Gobin-Rahimbux, N. Gooda Sahib-Kaudeer, N. Chooramun, S. Cheerkoot-Jalim, M. Heenaye-Mamode Khan, Z. Cadersaib, S. Kishnah and S. Elaheebocus	
The Economics of ‘Make in India’ Over ‘Buy (Import)’ Decision in Selected Technologies for the Indian Navy.....	345
Alok Bhagwat and Pradnya Vishwas Chitrap	

Naïve Bayes Classification on Student Placement Data: A Comparative Study of Data Mining Tools	363
Umang Mavani, Vivian Brian Lobo, Aditi Pednekar, Naomi Christianne Pereira, Rupesh Mishra and Nazneen Ansari	
Blockchain Technology: An Aid to the Governance of Smart Cities	373
Vishva Nitin Patel and Chhayaben Nitin Patel	
Machine Translation and Divergence Study for English–Maithili	383
Ritu Nidhi and Tanya Singh	
Detection of Liaison Between Health Outcome and Web Search	391
Sandip Roy and Amit Saha	
Wi-Fi-Based Portable Weather Station for Monitoring Temperature, Relative Humidity, Pressure, Precipitation, Wind Speed, and Direction	399
Indranil Sarkar, Bijoy Pal, Arnab Datta and Sandip Roy	
Automated Socio-psychological Knowledge Behavior Classified in E-Commerce Applying Various Machine Learning Techniques	405
P. Vijayaragavan, R. Ponnusamy and M. Arrmuthan	
ICI Reduction by Parallel Concatenated Encoder Using Wavelet Transforms	415
Noura Ali, M. I. Youssef and I. F. Tarrad	
Sentence-Level Emotion Detection from Text Based on Semantic Rules	423
Dibyendu Seal, Uttam K. Roy and Rohini Basak	
A Real-Time Automatic Human Facial Expression Recognition System Using Deep Neural Networks	431
Surekha Samsani and Vineel Abhinav Gottala	
Ballast Water Quality Compliance Monitoring Using IoT	443
K. Komathy	
Analysis of Malignancy Using Enhanced GraphCut-Based Clustering for Diagnosis of Bone Cancer	453
B. S. Vandana, P. J. Antony and Sathyavathi R. Alva	
Personalized News Media Extraction and Archival Framework with News Ordering and Localization	463
Shine K. George, V. P. Jagathy Raj and Santhosh Kumar Gopalan	
Analysis of Transfer Characteristics of Junctionless GaAs-Nanotube MOSFET with Hafnium Oxide Dielectric	473
Raj Kumar and Arvind Kumar	

Signal Integrity Analysis for Diameter-Dependent Mixed Carbon Nanotube Bundle Interconnects	481
Vijay Rao Kumbhare, Punya Prasanna Paltani and Manoj Kumar Majumder	
Assignment Tracking on Android Platform	491
Dhriti Chakraborty and Matangini Chattopadhyay	
Empowering Textile Crafts Through the Internet of Things Technology	501
Deepshikha and Pradeep Yammiyavar	
RSA Using Montgomery Powering Ladder on Dual Core	513
Pawan Singh, Ashwani Kumar and Mohit Kumar	
Validation and Optimization of Image Compression Algorithms	521
Nikhilesh Joshi and Tanuja Sarode	
Temperature Control Using Sliding Mode Control: An Experimental Approach	531
Krupa Narwekar and V. A. Shah	
Simulation Platforms for Wireless Sensor Networks: How to Select?	539
Aarti Kochhar, Pardeep Kaur and Preeti	
Design of an Architectural Framework for the Implementation of eHealth in Mauritius	547
Sudha Cheerkoot-Jalim and Kanishka Gopal	
Sentiment Analysis on Online Product Reviews	559
Rajesh Bose, Raktim Kumar Dey, Sandip Roy and Debabrata Sarddar	
An Improved Approach to Background Removal Using Haar-Based Preprocessing for Phase Features	571
M. B. Veena and Meena Deshpande	
Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors	585
Preeti Tiwari and Piyush Shukla	
A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT	595
Ashwani Kumar	
Time Series Forecasting Using Double Exponential Smoothing for Predicting the Major Ambient Air Pollutants	603
Rajesh Bose, Raktim Kumar Dey, Sandip Roy and Debabrata Sarddar	

Compression for DNA Sequences Using Huffman Encoding	615
Raju Bhukya, Subodh Yadav, Jitendra Kumar Sharma, Banabari Lal Sharma and Arun Kumar	
Use of Clustering Sentiments for Opinion Mining: An Experimental Analysis	625
D. Veeriah and J. Nageswara Rao	
ICT for HR in an Industrial Organisation for a Small but High-Impact Process (a Case Study of RINL, Visakhapatnam Steel Plant, India)	633
K. Satya Narayana, G. V. Ramesh and Debasish Ray	
Smart City and Challenges	643
Dinesh Kumar Saini and B. Y. Sandhiyaa	
Flow Shop Scheduling in Which Processing Time Connected with Probabilities and Job Delay Due to Maintenance for $M * N$ Machine	651
E. Janaki and A. Mohamed Ismail	
IoE-Enabled Smart Embedded System: An Innovative Way of Learning	659
Akshada Rathod, Prachi Ayare, Ramchandra Bobhate, Rajneeshkaur Sachdeo, Sambhaji Sarode and Jyoti Malhotra	
Multifactor Secure Login Using Graceful Labeled Paths	669
K. Vengata Krishnan and D. Lalitha	
Kolam Generated by Color Petri Nets	675
M. I. Mary Metilda and D. Lalitha	
Template Matching for Automatic Number Plate Recognition System with Optical Character Recognition	683
Arun Vaishnav and Manju Mandot	
An Exhaustive Study of Dominant Features for Natural Gas Consumption	695
Prabodh Kumar Pradhan, Sunil Dhal and Nilayam Kumar Kamila	
Visual Analytics Toward Prediction of Employee Erosion Through Data Science Tools	705
Priyanka Nair, Jaya Krishna and Devesh Kumar Srivastava	
Secured Forensic Framework for Various Users in the Virtualized Environment of Cloud	715
Gayatri S. Pandi and K. H. Wandra	
Document Classification Using Wikidata Properties	729
Sheeban Wasi, Madhurendra Sachan and Manuj Darbari	

Energy-Efficient Fuzzy-Logic-Based Data Aggregation in Wireless Sensor Networks	739
Sukhchandan Randhawa and Sushma Jain	
An Approach Toward More Accurate Forecasts of Air Pollution Levels Through Fog Computing and IoT	749
Bhavya Deep, Iti Mathur and Nisheeth Joshi	
Implementation of Dalal and Triggs Algorithm to Detect and Track Human and Non-Human Classifications by Using Histogram-Oriented Gradient Approach	759
Vijayalakshmi S. Katti, Sushitha S, Shweta Dhareshwar and K. Sowmya	
Safe Navigation for Elderly and Visually Impaired People Using Adhesive Tactile Walking Surface Indicators in Home Environment	771
Rajanala Vijaya Prakash and Srinath Taduri	
Extensible Attribute Similarity Data Mining for Categorical Data Streams in Web Usage Framework	779
N. Pushpalatha, S. Sai Satyanarayana Reddy and N. Subhash Chandra	
IAAS Service in the Public Domain: Impact of Various Security Components on Trust	789
Archana B. Saxena and Meenu Dave	
Serving the Dermatologists: Skin Diseases Detection	799
Savy Gulati and Rosepreet Kaur Bhogal	
A Green Computing Technique for Reducing Energy Consumption for Desktops Using Proxy Downloading	823
Meenakshi S. Arya, Ashwini Manjare, Payal Naik and Kamrul Haq Sheikh	
Enhancement of Brain MR-T1/T2 Images Using Mathematical Morphology	833
Anu Arya, Vikrant Bhatija, Mansi Nigam and Anuj Singh Bhaduria	
Author Index	841

About the Editors

Milan Tuba received his B.S. and M.S. in Mathematics, and M.S., M.Ph. and Ph.D. in Computer Science from the University of Belgrade and New York University. He was an Assistant Professor of Electrical Engineering at Cooper Union Graduate School of Engineering, New York. During that time, he was the founder and director of Microprocessor Lab and VLSI Lab, and led scientific projects. Since 1994, he has participated in the scientific projects for the Republic of Serbia's Ministry of Science. Currently he is the Vice Rector of Singidunum University, Serbia. His research interests include mathematical, queuing theory and heuristic optimizations applied to computer networks, image processing and combinatorial problems. He is the author of more than 150 scientific papers and a monograph. He has held ten plenary lectures at conferences at Harvard, Cambridge, Moscow, Paris and Istanbul. He is also a co-editor, member of the editorial board, scientific committee member and a reviewer for several international journals and conferences. He is a member of various international and scientific organizations: ACM, IEEE, AMS, SIAM and FNA.

Shyam Akashe is a Professor at ITM University, Gwalior, Madhya Pradesh, India. He completed his Ph.D. at Thapar University, Punjab and his M.Tech. in Electronics and Communication Engineering at the Institute of Technology & Management, Gwalior. He has authored 190 publications, including more than 50 papers in SCI indexed journals. His main research focus is low power system on chip (SoC) applications in which static random access memories (SRAMs) are omnipresent. He has authored two books entitled *Moore's Law Alive: Gate-All-Around (GAA) Next Generation Transistor* published by LAMBERT Academic Publishing, Germany and *Low Power High Speed CMOS Multiplexer Design* published by Nova Science Publishers, Inc., New York, USA. He has also published over 120 papers in leading national and international refereed journals of repute. Dr. Akashe has participated in numerous national and international conferences and presented over 100 papers.

Amit Joshi is a young entrepreneur and researcher who holds a B.Tech. in Information Technology and M.Tech. in Computer Science and Engineering. He pursued research in the areas of cloud computing and cryptography and has 6 years' academic and industrial experience in prestigious organizations in Udaipur and Ahmedabad. Currently, he is working as an Assistant Professor at the Department of Information Technology at Sabar Institute in Gujarat. He is an active member of ACM, CSI, AMIE, IACSIT-Singapore, IDES, ACEEE, NPA and many other professional societies. He is also Honorary Secretary of CSI Udaipur Chapter and Secretary of the ACM Udaipur Chapter. He has presented and published more than 30 papers in national and international journals as well as IEEE and ACM conferences. He has edited three books on diverse subjects including advances in open source mobile technologies, ICT for integrated rural development and ICT for competitive strategies. He has also organized more than 15 national and international conference and workshops including ICTCS 2014 in Udaipur and ACM-ICPS. He has received appreciation awards from The Institution of Engineers (India), ULC and from the SIG-WNs Computer Society of India for his contributions to society.

Analysis of Performance Measures of Computer Systems with Priority and Maximum Operation Time



Indeewar Kumar, Ashish Kumar and Monika Saini

Abstract In the present work, an effort has been used to analyze the profit behavior of the computer systems under the concepts of priority, maximum repair time, and preventive maintenance after a pre-specific maximum operation time. For this purpose, a stochastic model has been developed by giving priority to software upgradation jointly over hardware repair and replacement. And the profit function of the anticipated stochastic model has been compared with some already existing models. A single repair facility is available. SMP and RPT have been used to develop recurrence relations for various performance measures. Graphical representation of the profit difference has been made to highlight the importance of the study.

Keywords Computer system · Economic analysis · Priority · Preventive maintenance · Maximum operation time

1 Introduction

Many studies have been carried out in the field of computer system's reliability. Most of the researchers concentrate either on hardware reliability or software reliability. Friedman and Tran [1], Welke et al. [2] and Lai et al. [3] mainly focused on the analysis of hardware/software reliability of computer system. Anand et al. [4] developed first time a stochastic model by considering computer system as a single entity a combination of hardware and software with independent hardware and software failures. Kumar et al. [5] made an effort to develop a reliability model with maximum repair and preventive maintenance after a pre-specific maximum operation time. Kumar and Malik [6, 7] used the concept of priority to s/w upgradation over h/w repair and h/w replacement to develop the reliability model for computer systems. But, the concept of priority to software upgradation over hardware repair and hardware

I. Kumar · A. Kumar (✉) · M. Saini

Department of Mathematics & Statistics, Manipal University Jaipur, Jaipur 303007, Rajasthan, India

e-mail: ashishbarak2020@gmail.com

replacement has not been discussed so far by the researchers jointly. Therefore, to strengthen the existing literature, the present article is designed with an object to analyze a computer system economically using the concept of priority to s/w upgradation over h/w repair activities (repair and replacement) jointly. Kumar and Saini [8] analyzed comparatively some measures of a computer system using the concept of priority. The profit analysis of a computing device under priority and s/w rejuvenation has been carried out by Kumar et al. [9]. Kumar et al. [10] studied various probabilistically performance measures of repairable cold standby systems under Weibull failure and repair laws. For this purpose, a stochastic model has been developed by giving priority to software upgradation jointly over hardware repair and replacement. And the profit of the proposed model has been compared with some already existing models. A single repair facility is available. SMP and RPT have been used to develop recurrence relations for various performance measures. To show the importance of the present work, graphs with respect to preventive maintenance rate have been drawn for profit difference of existing model with Kumar and Malik [7] and Kumar et al. [5].

2 Notations and Model Description

By using the same set of assumptions and notations proposed by Kumar and Malik [6], the possible transition states of the proposed models have been obtained and appended below:

Operative States:

$$\begin{aligned} S_0 &= (N_o, Cs), \quad S_1 = (N_o, Pm), \quad S_2 = (N_o, HFur), \\ S_3 &= (N_o, SFurp), \quad S_4 = (N_o, HFurp), \end{aligned}$$

Failed States:

$$\begin{aligned} S_5 &= (HFUR, Wpm), \quad S_6 = (HFwr, PM), \quad S_7 = (SFURP, HFwr), \\ S_8 &= (PM, SFwrp), \quad S_9 = (SFURP, WPm), \quad S_{10} = (SFURP, SFwrp), \\ S_{11} &= (HFwr, SFurp), \quad S_{12} = (HFUR, HFwr), \quad S_{13} = (WPm, (WPm, PM)), \\ S_{14} &= (HFurp, HFWR), \quad S_{15} = (HFurp, WPM), \quad S_{16} = (HFURP, Wpm), \\ S_{17} &= (HFwrp, SFurp), \quad S_{18} = (HFURP, HFwr) \end{aligned}$$

Regenerative States:

$$\begin{aligned} S_0 &= (N_o, Cs), \quad S_1 = (N_o, Pm), \quad S_2 = (N_o, HFur), \\ S_3 &= (N_o, SFurp), \quad S_4 = (N_o, HFurp), \quad S_{11} = (HFwr, SFurp), \\ S_{17} &= (HFwrp, SFurp) \end{aligned}$$

3 Transition Probabilities of the Anticipated Model

By using the rules of probability theory and t.p.m., Equation (1) gives the transition probabilities at all states

$$\begin{aligned}
 p_{ij} = Q_{ij}(\infty) &= \int_0^{\infty} q_{ij}(t) dt \text{ as} & (1) \\
 p_{01} &= \frac{\alpha_0}{A}, \quad p_{02} = \frac{a\lambda_1}{A}, \quad p_{03} = \frac{b\lambda_2}{A}, \quad p_{10} = f^*(A), \\
 p_{16} &= \frac{a\lambda_1}{A}[1 - f^*(A)] = p_{12.6}, \quad p_{18} = \frac{b\lambda_2}{A}[1 - f^*(A)] = p_{13.8}, \\
 p_{1.13} &= \frac{\alpha_0}{A}[1 - f^*(A)] = p_{11.13}, \quad p_{20} = g^*(B), \\
 p_{24} &= \frac{\beta_0}{B}[1 - g^*(B)], \\
 p_{25} &= \frac{\alpha_0}{B}[1 - g^*(B)], \quad p_{2.11} = \frac{b\lambda_2}{B}[1 - g^*(B)], \\
 p_{2.12} &= \frac{a\lambda_1}{B}[1 - g^*(B)], \quad p_{30} = h^*(A), \\
 p_{37} &= \frac{a\lambda_1}{A}[1 - h^*(A)] = p_{32.7}, \\
 p_{39} &= \frac{\alpha_0}{A}[1 - h^*(A)] = p_{3.1.9}, \quad p_{40} = m^*(A), \\
 p_{3.10} &= \frac{b\lambda_2}{A}[1 - h^*(A)] = p_{33.10}, \quad p_{51} = g^*(\beta_0), \\
 p_{5.16} &= 1 - g^*(\beta_0), \quad p_{4.16} = \frac{\alpha_0}{A}[1 - m^*(A)] = p_{4.1.16}, \\
 p_{62} &= f^*(0), \quad p_{72} = h^*(0), \quad p_{83} = f^*(0), \quad p_{91} = h^*(0), \\
 p_{10.3} &= h^*(0), \quad p_{11.2} = h^*(\beta_0), \quad p_{4.17} = \frac{b\lambda_2}{A}[1 - m^*(A)], \\
 p_{12.2} &= g^*(\beta_0), \quad p_{12.14} = 1 - g^*(\beta_0), \quad p_{13.1} = f^*(0), \\
 p_{4.19} &= \frac{a\lambda_1}{A}[1 - m^*(A)] = p_{42.18}, \quad p_{15.2} = m^*(0), \\
 p_{16.1} &= m^*(0), \quad p_{17.4} = h^*(0), \quad p_{18.2} = m^*(0), \\
 p_{21.5} &= \frac{\alpha_0}{B}[1 - g^*(B)]g^*(\beta_0), \\
 p_{21.5.15.} &= \frac{\alpha_0}{B}[1 - g^*(B)][1 - g^*(\beta_0)], \\
 p_{2.11} &= \frac{b\lambda_2}{B}[1 - g^*(B)], \quad p_{22.12} = \frac{a\lambda_1}{B}[1 - g^*(B)]g^*(\beta_0),
 \end{aligned}$$

$$p_{22.12,14} = \frac{a\lambda_1}{B} [1 - g^*(B)] [1 - g^*(\beta_0)],$$

where $A = a\lambda_1 + b\lambda_2 + \alpha_0$ and $B = a\lambda_1 + b\lambda_2 + \alpha_0 + \beta_0$ (2)

4 Mean Sojourn Times of the Anticipated Model

$$\begin{aligned}\mu_0 &= \frac{1}{a\lambda_1 + b\lambda_2 + \alpha_0}, & \mu_1 &= \frac{1}{a\lambda_1 + b\lambda_2 + \alpha_0 + \alpha}, & \mu_2 &= \frac{1}{a\lambda_1 + b\lambda_2 + \alpha_0 + \theta + \beta_0}, \\ \mu_3 &= \frac{1}{a\lambda_1 + b\lambda_2 + \alpha_0 + \beta}, & \mu_4 &= \frac{1}{a\lambda_1 + b\lambda_2 + \alpha_0 + \gamma},\end{aligned}\quad (3)$$

5 Steady State Availability

By using simple probabilistic arguments and

$$\begin{aligned}M_0(t) &= e^{-(a\lambda_1 + b\lambda_2 + \alpha_0)t}, & M_1(t) &= e^{-(a\lambda_1 + b\lambda_2 + \alpha_0)t} \overline{F(t)} \\ M_2(t) &= e^{-(a\lambda_1 + b\lambda_2 + \alpha_0 + \beta_0)t} \overline{G(t)}, & M_3(t) &= e^{-(a\lambda_1 + b\lambda_2 + \alpha_0)t} \overline{H(t)} \\ M_4(t) &= e^{-(a\lambda_1 + b\lambda_2 + \alpha_0)t} \overline{M(t)}\end{aligned}\quad (4)$$

The recurrence relations for $A_i(t)$, where $A_i(t)$ is the probability of upstate at time instant t , have been obtained as follows:

$$A_i(t) = M_i(t) + \sum_j q_{i,j}^{(n)}(t) \odot A_j(t) \quad (5)$$

Applying Laplace transformation on Eq. (5) and solving for $A_0^*(s)$. The availability is given by

$$A_0(\infty) = \lim_{s \rightarrow 0} s A_0^*(s) = \frac{N_2}{D_2}, \quad (6)$$

where

$$\begin{aligned}N_2 = & (-p_{24})\{\mu_0[(1 - p_{11.13})p_{42.18}(1 - p_{33.10}) + (1 - p_{33.10})p_{41.16}p_{12.6} \\ & - p_{13.8}\{p_{42.18}p_{31.9} - p_{41.16}p_{32.7}\} + \mu_1[p_{01}p_{42.18}(1 - p_{33.10}) - p_{02} \\ & p_{41.16}(1 - p_{33.10}) + p_{03}\{p_{42.18}p_{31.9} - p_{41.16}p_{32.7}\}] + \mu_3[p_{01}p_{42.18} \\ & p_{13.8} - p_{02}p_{41.16}p_{13.8} + p_{03}\{p_{42.18}(1 - p_{11.13}) + p_{41.16}p_{12.6}\}] - \mu_4[p_{01} \\ & \{(1 - p_{33.10})p_{12.6} + p_{32.7}p_{13.8}\} + p_{01}\{(1 - p_{11.13})(1 - p_{33.10}) \\ & - p_{13.8}p_{31.9}\} + p_{03}\{(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}\}]\} + (1 - p_{4.17}p_{17.4})\{\mu_0\end{aligned}$$

$$\begin{aligned}
& [(1 - p_{11.13})(1 - p_{33.10})(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) - (p_{22.5} + \\
& p_{21.5.15})(1 - p_{33.10})p_{12.6} - p_{13.8}\{(1 - p_{22.12} - p_{22.12.14} \\
& - p_{2,11}p_{11.2})p_{31.9} + (p_{22.5} + p_{21.5.15})p_{32.7}\}] + \mu_1[p_{01}(1 - p_{33.10})(1 - \\
& p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) + (p_{22.5} + p_{21.5.15})(1 - p_{33.10})p_{02} \\
& + p_{03}\{(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2})p_{31.9} + (p_{22.5} + p_{21.5.15})p_{32.7}\}] \\
& + \mu_2[p_{01}\{(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}\} + p_{02}\{(1 - p_{11.13})(1 - p_{33.10}) \\
& - p_{13.8}p_{31.9}\} + p_{03}\{(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}\} + \mu_4[p_{01}p_{13.8}(1 - \\
& p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) + (p_{22.5} + p_{21.5.15})p_{13.8}p_{02} + p_{03}\{(1 - \\
& p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2})(1 - p_{11.13}) - (p_{22.5} + p_{21.5.15})p_{12.6}\}]] \\
\end{aligned}$$

and

$$\begin{aligned}
D_2 = & (-p_{24})\{\mu_0[(1 - p_{11.13})p_{42.18}(1 - p_{33.10}) + (1 - p_{33.10})p_{41.16}p_{12.6} \\
& - p_{13.8}\{p_{42.18}p_{31.9} - p_{41.16}p_{32.7}\} + \mu'_1[p_{01}p_{42.18}(1 - p_{33.10}) \\
& - p_{02}p_{41.16}(1 - p_{33.10}) + p_{03}\{p_{42.18}p_{31.9} - p_{41.16}p_{32.7}\}] \\
& + \mu'_3[p_{01}p_{42.18}p_{13.8} - p_{02}p_{41.16}p_{13.8} + p_{03}\{p_{42.18}(1 - p_{11.13}) + p_{41.16}p_{12.6}\}] \\
& - (\mu'_4 + p_{4.17}\mu_{17})[p_{01}\{(1 - p_{33.10})p_{12.6} + p_{32.7}p_{13.8}\} \\
& + p_{01}\{(1 - p_{11.13})(1 - p_{33.10}) - p_{13.8}p_{31.9}\} \\
& + p_{03}\{(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}\}] \\
& + (1 - p_{4.17}p_{17.4})\{\mu_0[(1 - p_{11.13})(1 - p_{33.10}) \\
& (1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) - (p_{22.5} + p_{21.5.15})(1 - p_{33.10})p_{12.6} \\
& - p_{13.8}\{(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2})p_{31.9} + (p_{22.5} + p_{21.5.15})p_{32.7}\}] \\
& + \mu'_1[p_{01}(1 - p_{33.10})(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) \\
& + (p_{22.5} + p_{21.5.15})(1 - p_{33.10})p_{02} \\
& + p_{03}\{(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2})p_{31.9} + (p_{22.5} + p_{21.5.15})p_{32.7}\}] \\
& + (\mu'_2 + p_{2,11}\mu_{11})[p_{01}\{(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}\} \\
& + p_{02}\{(1 - p_{11.13})(1 - p_{33.10}) - p_{13.8}p_{31.9}\} \\
& + p_{03}\{(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}\}] \\
& + \mu'_3[p_{01}p_{13.8}(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2}) + (p_{22.5} + p_{21.5.15})p_{13.8}p_{02} \\
& + p_{03}\{(1 - p_{22.12} - p_{22.12.14} - p_{2,11}p_{11.2})(1 - p_{11.13}) \\
& - (p_{22.5} + p_{21.5.15})p_{12.6}\}]\} \\
\end{aligned}$$

6 Busy Period Analysis for Server

By using simple probabilistic arguments and

$$\begin{aligned}
W_1 &= e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \bar{F}(t) + (\alpha_0 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{F}(t) \\
&\quad + (a\lambda_1 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{F}(t) + (b\lambda_2 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{F}(t) \\
W_2 &= e^{-(a\lambda_1+b\lambda_2+\alpha_0+\beta_0)t} \bar{G}(t) + (\alpha_0 e^{-(a\lambda_1+b\lambda_2+\alpha_0+\beta_0)t} \odot 1) \bar{G}(t) \\
&\quad + (a\lambda_1 e^{-(a\lambda_1+b\lambda_2+\alpha_0+\beta_0)t} \odot 1) \bar{G}(t) + (b\lambda_2 e^{-(a\lambda_1+b\lambda_2+\alpha_0+\beta_0)t} \odot 1) \bar{G}(t) \\
W_3 &= e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \bar{H}(t) + (\alpha_0 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{H}(t) \\
&\quad + (a\lambda_1 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{H}(t) + (b\lambda_2 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{H}(t) \\
W_4 &= e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \bar{M}(t) + (\alpha_0 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{M}(t) + (a\lambda_1 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \\
&\quad \bar{M}(t) + (b\lambda_2 e^{-(a\lambda_1+b\lambda_2+\alpha_0)t} \odot 1) \bar{M}(t), \quad W_{17} = \bar{H}(t), \quad W_{11} = \bar{H}(t)
\end{aligned}$$

The recurrence relations for $B_i^P(t)B_i^R(t)B_i^S(t)$ and $B_i^{HRP}(t)$, where $B_i(t)$ is the probability that server is busy in various repair activities at time instant t , have been obtained as follows:

$$\begin{aligned}
B_i^P(t) &= W_i(t) + \sum_j q_{i,j}^{(n)}(t) \odot B_j^P(t) \\
B_i^R(t) &= W_i(t) + \sum_j q_{i,j}^{(n)}(t) \odot B_j^R(t) \\
B_i^S(t) &= W_i(t) + \sum_j q_{i,j}^{(n)}(t) \odot B_j^S(t) \text{ and} \\
B_i^{HRP}(t) &= W_i(t) + \sum_j q_{i,j}^{(n)}(t) \odot B_j^{HRP}(t)
\end{aligned} \tag{7}$$

Applying Laplace transformation on Eq. (7) and solving for $B_i^{*P}(s)B_i^{*R}(s)B_i^{*S}(s)$ and $B_i^{*HRP}(s)$, the time for which server is busy due to repair and replacements, respectively, is given by

$$\begin{aligned}
B_0^H &= \lim_{s \rightarrow 0} s B_0^{*H}(s) = \frac{N_3^H}{D_2}, \quad B_0^S = \lim_{s \rightarrow 0} s B_0^{*S}(s) = \frac{N_3^S}{D_2}, \\
B_0^R &= \lim_{s \rightarrow 0} s B_0^{*R}(s) = \frac{N_3^R}{D_2} \text{ And} \\
B_0^{HRP} &= \lim_{s \rightarrow 0} s B_0^{*HRP}(s) = \frac{N_3^{HRP}}{D_2}
\end{aligned} \tag{8}$$

where

$$\begin{aligned}
N_3^P &= W_1^*(0)\{p_{24}[-p_{01}p_{42.18}p_{13.8} + p_{02}p_{41.16}p_{13.8} - p_{03}\{p_{31.9}p_{42.18} - p_{41.16}p_{32.7}\}] + \\
&\quad (1 - p_{4.17}p_{17.4})[p_{01}(1 - p_{33.10})(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2}) + (p_{22.5} + p_{21.5.15}) \\
&\quad (1 - p_{33.10})p_{02} + p_{03}\{(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2})p_{31.9} + (p_{21.5} + p_{21.5.15})p_{32.7}\}] \\
N_3^R &= W_2^*(0)(1 - p_{4.17}p_{17.4})\{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] \\
&\quad + p_{02}[(1 - p_{33.10})(1 - p_{11.13}) - p_{13.8}p_{31.9}]
\end{aligned}$$

$$\begin{aligned}
& + p_{03}[(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}]\} \\
N_3^S &= (p_{24})[-W_3^*(0)\{p_{01}p_{42.18}p_{13.8} - p_{02}p_{41.16}p_{13.8} \\
& + p_{03}[(1 - p_{11.13})p_{42.18} - p_{41.16}p_{12.6}]\} + p_{4.17}W_{11}^*(0) \\
& \{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] + p_{02}[(1 - p_{33.10})(1 - p_{11.13}) - p_{13.8}p_{31.9}] \\
& + p_{03}[(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}]\} \\
& + (1 - p_{4.17}p_{17.4})[p_{24}W_{11}^*(0)\{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] \\
& + p_{02}[(1 - p_{33.10})(1 - p_{11.13}) - p_{13.8}p_{31.9}] \\
& + p_{03}[(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}]\} + W_3^*(0) \\
& \{p_{01}p_{13.8}(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2}) + (p_{22.5} + p_{21.5.15})p_{13.8}p_{02} \\
& + p_{03}\{(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2})(1 - p_{11.13}) - (p_{21.5} + p_{21.5.15})p_{12.6}\}\}] \\
N_3^{HRP} &= W_4^*p_{24}[p_{01}\{(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}\} + p_{02}\{(1 - p_{33.10})(1 - p_{11.13}) \\
& - p_{13.8}p_{31.9}\} + p_{03}\{(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}\}]
\end{aligned}$$

and D_2 is already mentioned.

7 Expected Number of Replacements (ENR) of the Units

Let $R_i^H(t)$ and $R_i^S(t)$ denote the expected number of replacement of the failed hardware and failed software components by the server in $(0, t]$ given that the system entered the regenerative state i at $t = 0$. The recursive relations for $R_i^H(t)$ and $R_i^S(t)$ are given as

$$\begin{aligned}
R_i^H(t) &= \sum_j Q_{i,j}^{(n)}(t) \circledast [\delta_j + R_j^H(t)], \\
R_i^S(t) &= \sum_j Q_{i,j}^{(n)}(t) \circledast [\delta_j + R_j^S(t)] \\
N_i(t) &= \sum_j Q_{i,j}^{(n)}(t) \circledast [\delta_j + N_j(t)]
\end{aligned} \tag{9}$$

where j is any regenerative state to which the given regenerative state i transits and

$$\delta_j = \begin{cases} 1, & \text{if } j \in \text{regenerative state} \\ 0 & \text{if } j \notin \text{regenerative state} \end{cases}$$

Taking Laplace Stieltjes transformation of relations (9) and, solving for $\tilde{R}_0^H(s)$ and $\tilde{R}_0^S(s)$. The expression (10) gives the ENR per unit time to the hardware and software failures, respectively.

$$\begin{aligned}
R_0^H(\infty) &= \lim_{s \rightarrow 0} s \tilde{R}_0^H(s) = \frac{N_4^H}{D_2}, \quad R_0^S(\infty) = \lim_{s \rightarrow 0} s \tilde{R}_0^S(s) \\
& = \frac{N_4^S}{D_2} \quad \text{and} \quad N_0(\infty) = \lim_{s \rightarrow 0} s \tilde{N}_0(s) = \frac{N_5}{D_2}
\end{aligned} \tag{10}$$

$$\begin{aligned}
N_4^H = & (p_{40} + p_{41.16} + p_{42.18})(p_{24})\{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] + p_{02} \\
& [(1 - p_{11.13})(1 - p_{33.10}) - p_{31.9}p_{13.8}] + p_{03}[(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}]\} \\
& + (p_{22.12.14} + p_{21.5.15})(1 - p_{17.4}p_{4.17})\{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] \\
& + p_{02}[(1 - p_{11.13})(1 - p_{33.10}) - p_{31.9}p_{13.8}] \\
& + p_{03}[(1 - p_{11.13})p_{32.7} + p_{31.9}p_{12.6}]\}
\end{aligned}$$

$$\begin{aligned}
N_4^S = & (-p_{24})\{[p_{01}p_{42.18}p_{13.8} - p_{02}p_{41.16}p_{13.8} \\
& + p_{03}\{(1 - p_{11.13})p_{42.18} - p_{41.16}p_{12.6}\}] - p_{17.4}p_{4.17} \\
& \{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] \\
& + p_{02}[(1 - p_{33.10})(1 - p_{11.13}) - p_{13.8}p_{31.9}] \\
& + p_{03}[(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}]\} \\
& + (1 - p_{4.17}p_{17.4})[p_{2.11}p_{11.2}\{p_{01}[(1 - p_{33.10})p_{12.6} + p_{13.8}p_{32.7}] \\
& + p_{02}[(1 - p_{33.10})(1 - p_{11.13}) \\
& - p_{13.8}p_{31.9}] + p_{03}[(1 - p_{11.13})p_{32.7} + p_{12.6}p_{31.9}]\}] \\
& + p_{01}p_{13.8}\{p_{01}p_{13.8}(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2}) \\
& + (p_{22.5} + p_{21.5.15})p_{13.8}p_{02} + p_{03} \\
& \{(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2})(1 - p_{11.13}) - (p_{21.5} + p_{21.5.15})p_{12.6}\}\}] \\
N_5 = & (-p_{24}) \left[\begin{array}{l} (1 - p_{11.13})p_{42.18}(1 - p_{33.10}) + (1 - p_{33.10})p_{41.16}p_{12.6} \\
- p_{13.8} \left\{ \begin{array}{l} p_{42.18}p_{31.9} - \\ p_{41.16}p_{32.7} \end{array} \right\} + (1 - p_{4.17}p_{17.4}) \end{array} \right] \\
& (1 - p_{11.13})(1 - p_{33.10})(1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2}) \\
& -(p_{22.5} + p_{21.5.15})(1 - p_{33.10})p_{12.6} \\
& - p_{13.8} \{ (1 - p_{22.12} - p_{22.12.14} - p_{2.11}p_{11.2})p_{31.9} + (p_{22.5} + p_{21.5.15})p_{32.7} \}]
\end{aligned}$$

where D_2 is already mentioned.

8 Profit Analysis

The profit generated by system model in the long run has been obtained by using relation (11) as follows:

$$P = K_0 A_0 - K_1 B_0^P - K_2 B_0^R - K_3 B_0^S - K_4 B_0^{HRP} - K_5 R_0^H - K_6 R_0^S - K_7 N_0 \quad (11)$$

K_0 = Revenue per unit uptime of the system

K_i ($i = 1, 2, 3, 4, 5, 6, 7$) = Cost per unit time for which performs various duties.

9 Comparative Analysis

Profit comparison of the present model with the model of Kumar et al. [5]:

The present model is more profitable over the model discussed in Kumar et al. [5]. Hence, the concept of priority to s/w upgradation over h/w repair activities is more profitable over the system in which no priority is given. The graphical behavior of the profit difference of these models has been shown in Fig. 1 with respect to PM rate (α) and for constant values of other parameters.

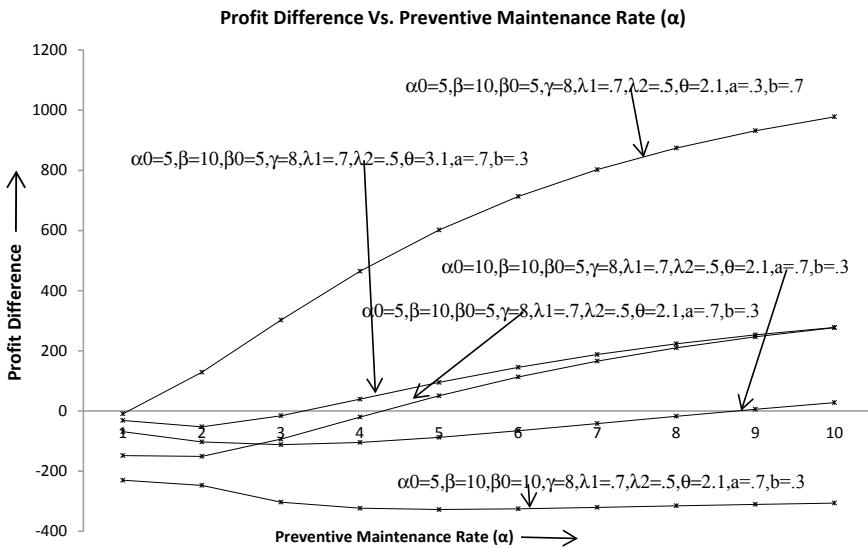


Fig. 1 Kumar et al. [5] versus proposed model

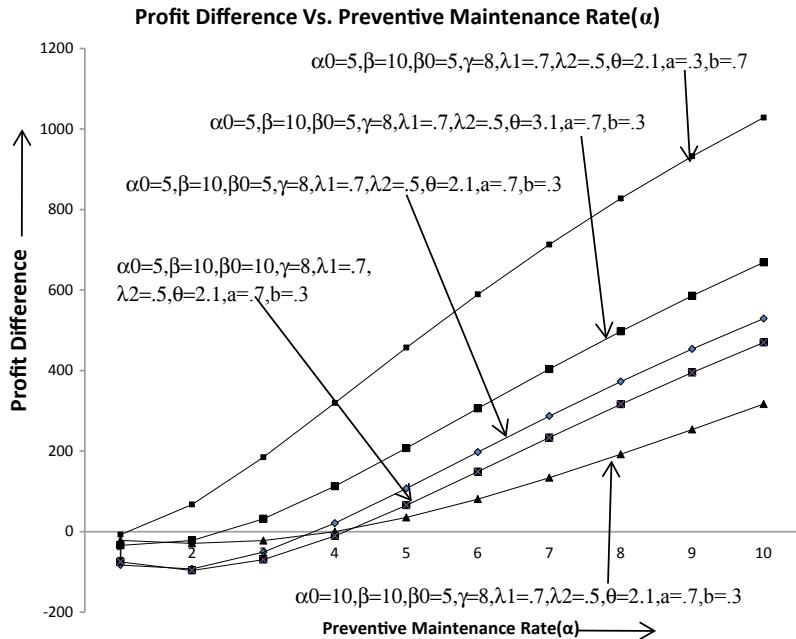


Fig. 2 Kumar and Malik [6] versus proposed model

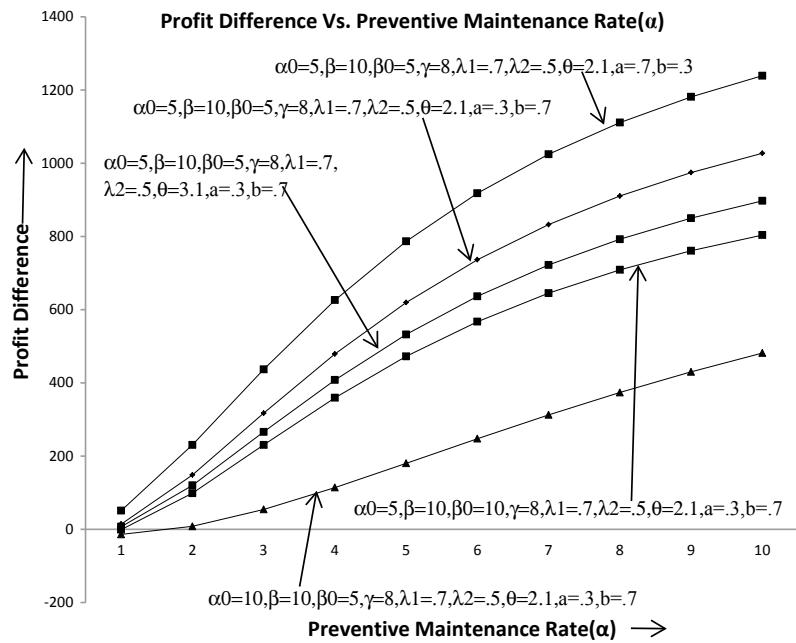


Fig. 3 Kumar and Malik [7] versus proposed model

Profit comparison the present model with the models of Kumar and Malik [6] and Kumar and Malik [7]:

The present model has more profit as compared to Kumar and Malik [6] while in comparison with Kumar and Malik [7], the present model has less profit for $\alpha < 3$. However, the present model is more profitable for $\alpha > 3$. Thus, the work reveals that a system in which priority is given to s/w upgradation jointly over h/w repair and replacement will have more profit as compared to the system where such a priority is given over h/w repair and replacement separately provided that the preventive maintenance rate (α) > 3. The behavior of profit difference of these models has also been shown in Figs. 2 and 3 with respect to PM rate (α) and for constant values of other parameters.

References

1. Friedman, M.A., Tran, P.: Reliability techniques for combined hardware/software systems. In: Proceedings Annual Reliability and Maintainability Symposium, pp. 290–293. IEEE (1992)
2. Welke, S.R., Johnson, B.W., Aylor, J.H.: Reliability modeling of hardware/software systems. *IEEE Trans. Reliab.* **44**(3), 413–418 (1995)
3. Lai, C.D., Xie, M., Poh, K.L., Dai, Y.S., Yang, P.: A model for availability analysis of distributed software/hardware systems. *Inf. Softw. Technol.* **44**, 343–350 (2002)
4. Anand, J., Kumar, A., Malik, S.C.: Cost-analysis of a computer system with hardware repair subject to inspection and arbitrary distributions for hardware and software replacement. *Int. J. Eng. Sci. Technol.* **80**(5), 693–709 (2011)
5. Kumar, A., Malik, S.C., Barak, M.S.: Reliability Modeling of a computer system with independent H/W and S/W failures subject to maximum operation and repair times. *Int. J. Math. Achiev.* **3**(7), 2622–2630 (2012)
6. Kumar, A., Malik, S.C.: Profit analysis of a computer system with priority to software replacement over hardware repair subject to maximum operation and repair times. *Int. J. Eng. Sci. Technol.* **3**(10), 7452–7468 (2011)
7. Kumar, A., Malik, S.C.: Profit analysis of a computer system with priority to software replacement over hardware repair subject to maximum operation and repair times. *Int. J. Pure Appl. Math.* **80**(5), 693–709 (2012)
8. Kumar, A., Saini, M.: Comparison of various reliability measures of a computer system with the provision of priority. In: Afzalpulkar, N., Srivastava, V., Singh, G., Bhatnagar, D. (eds.) *Proceedings of the International Conference on Recent Cognizance in Wireless Communication & Image Processing*. Springer, New Delhi (2016)
9. Kumar A., Saini M., Srivastava D.K.: Profit analysis of a computing machine with priority and s/w rejuvenation. In: Mishra, D., Nayak, M., Joshi, A. (eds.) *Information and Communication Technology for Sustainable Development. Lecture Notes in Networks and Systems*, vol. 9. Springer, Singapore (2017)
10. Kumar, I., Kumar, A., Saini, M., Devi, K.: Probabilistic analysis of performance measures of redundant systems under Weibull failure and repair laws. In: Vishwakarma H., Akashe, S. (eds.) *Computing and network sustainability. Lecture Notes in Networks and Systems*, vol 12. Springer, Singapore (2017)

Novel Concept of Spelling Correction for Semantic Tourism Search Interface



Shilpa Laddha and Pradip M. Jawandhiya

Abstract In this technical era, searching the Web has accelerated, but the user satisfaction is still not up to the mark. However, retrieval of required data is yet a challenging job. Users still struggle to get the precise facts on the Internet. To fetch the most correct end result is a hard mission due to the giant unstructured repository on the Web, huge list of results returned by search engines, etc. Spelling errors in query are the main attribute that degrades the performances of any search interface. In this paper, the effectiveness of the search interface is improved and the performance is accelerated by implementing novel spelling correction module which is implemented and tested on Indian tourism semantic search interface to handle misspelled words in query.

Keywords Information retrieval · Semantic search engine · Spelling correction · Tourism

1 Introduction

In this Internet era, existence without World Wide Web (WWW) is beyond imagination which is the quickest, simplest, and most practical medium of correspondence. With the lightning advancement of WWW, Web search tools [1, 2] have turned into the fundamental device of information retrieval to look through the Web information. Usually, user enters some keywords in the browser, and search engine [3, 4] performs the query using keywords and provides relevant results as output. However, retrieval of required data is yet a challenging job. To get the most correct end result for the searched question is a hard mission due to the giant unstructured repository

S. Laddha ()

Government College of Engineering, Aurangabad 431001, India

e-mail: kabrageca@gmail.com

P. M. Jawandhiya

P. L. Institute of Technology and Management, Buldana 443001, India

e-mail: pmjawandhiya@rediffmail.com

on the Web, and misspelled keywords in query [5] result in huge number of irrelevant search results. In this paper, the efficiency of the search engine is enhanced by implementing novel spelling correction module to handle misspelled words in query, thereby accelerating the performance of semantic search interface. Within the scope of this research, Indian Tourism [6, 7] is chosen as a domain of study and the performance is evaluated [8, 9]. Sections 2, 3 and 4 are the core of the paper presenting the challenges, system architecture, and performance analysis. Section 5 concludes the paper.

2 Challenges

Spelling errors are the main attribute that degrades the performances of any search interface. Nagata et al. [10] describe a method specially designed for automatically correcting grammatical and spelling errors using spelling error correction models from raw learner corpora which outperform the previous edit-distance-based and language-model-based methods. While entering the query in the semantic tourism search interface, if the user enters wrong spelling of a city or state, the system is unable to match the misspelled city or state name in query with available valid city or state names. The system is unable to interpret the query and thereby unable to provide any result. To resolve this, novel spelling correction module is designed which corrects the misspelled city or state name with the closest match in the dictionary containing a list of valid state and city names.

3 Architecture of Spelling Correction

The architecture of spelling correction is shown in Figs. 1 and 2.

3.1 Working of Spelling Correction Module

- (i) Query prototype manager as discussed in [11–13] holds all query prototypes and respective tourism service. The new query prototype is framed by removing the () of ontological token.
- (ii) Prototype mapper [14] is designed to match input query with new query prototype obtained in step 1.
- (iii) New query prototype is further updated by replacing each template token [12, 13] by regular expression [a-zA-Z][\s]+. For identifying the template token, new query prototype is split into elements by checking whether the starting index of element is “[” and ending index is “]”; if this condition is satisfied, then

Fig. 1 Spelling correction algorithm on dictionary

Spelling Correction Algorithm on Dictionary

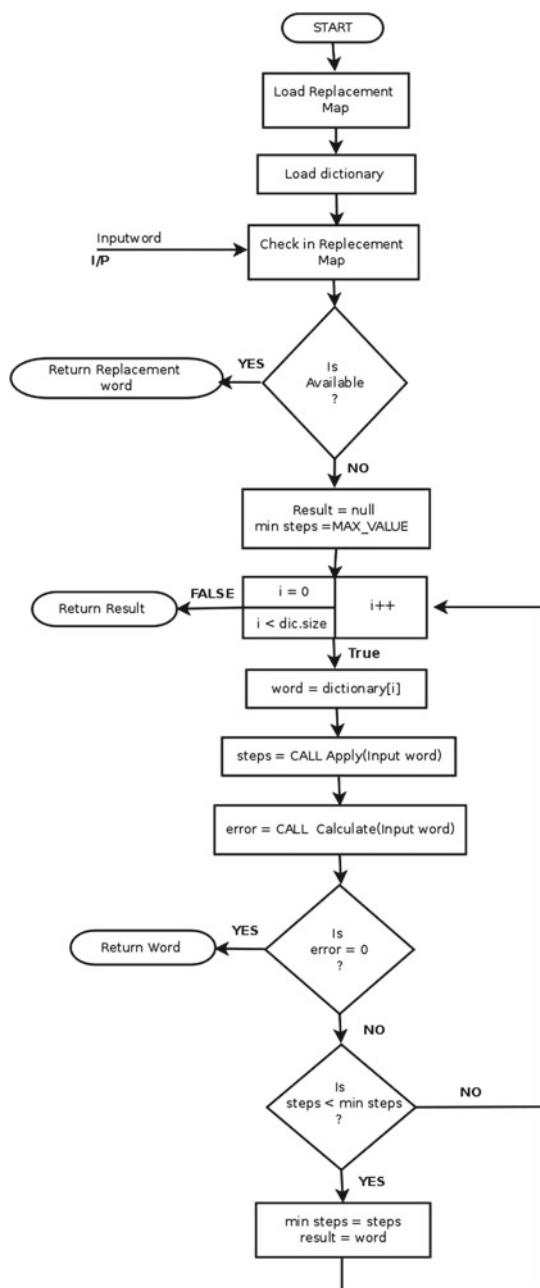
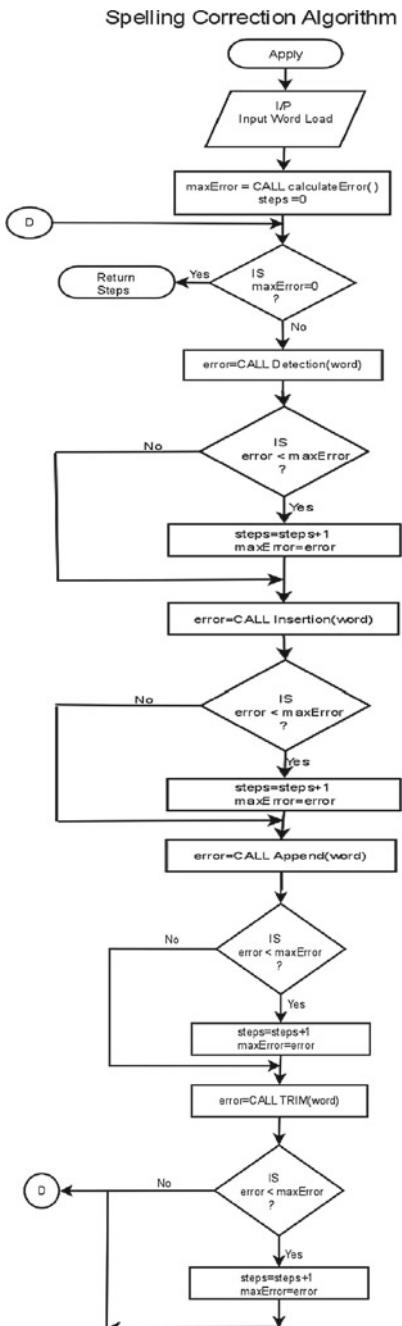


Fig. 2 Architecture of spelling correction



- element in new query prototype is replaced by regular expression [a-z\\s]+. New query prototype= distance from [a-z\\s]+ to [a-z\\s]+.
- (iv) These new query prototype and user query are given to Pattern Mapper to match the pattern of input query with new query prototype. If the match is found, it returns true; else, it returns false.
 - (v) The TemplateValueFinder is designed to get the value of the template tokens [fromcity]\[tocity]\[city]\[state] from user query. The prototype is trimmed by removing extra spaces. The FindTemplateValue is designed to get left string of query prototype before [] and right string of query prototype after []. For the new query prototype distance from [fromcity] to [tocity]: For template token 1 [fromcity]: Left string becomes “distance from,” and right string becomes “to.” For template token 2 [tocity]: Left string becomes “to,” and right string becomes “null.”
 - (vi) The check is made whether left string of new query prototype matches with query; if it is matched, then retrieve string at current index and store in [fromcity]. Similarly, the check is made whether right string of query prototype matches with query; if it is matched, then retrieve string at current index and store it in [tocity]. These fromcity and tocity were input word for the spelling correction algorithm.
 - (vii) Novel spelling correction algorithm is designed and implemented in the following way. E.g., input word=abcd Most nearest available word=abcde then this technique change i/p “abcd” to nearest available word in dictionary say “abcde.”
 - (viii) Dictionary of available states and cities names is maintained. If state\city name consists of stop word, then stop word is removed, and the words are stored as a separate entry in the dictionary (Andaman and Nicobar). Two separate entries are there for Andaman and Nicobar. Procedure apply (inputWord, availableWord) in spelling correction module is invoked, and a number of steps are evaluated.
 - (ix) The misspelled word is given to the spelling correction algorithm along with the available words in dictionary, and the total error is calculated by matching input word and the available word as follows. The length of input word=length1 and length of available word=length2. The loop is iterated for min (length1, length2) times. The characters at each position of input and available words are compared. If they are not matched, then total error is incremented. At the end, total error calculated is returned. If the error is zero, then process is terminated; else, the different methods like delete(), insert(), replace(), append(), trim() are invoked. The detailed working of different methods is discussed below.
 - (x) delete(inputWords, availableWord) It stores word in data object, and the check is made for mismatch position. Length of the input word and available word is evaluated and stored in length1 and length2, respectively, and the loop is iterated for min (length1,length2) times. During each iteration, character at current index in both words is compared, and if they are same, then -1 is returned; else, the mismatch index is returned to delete () method for deletion,

data is set to new word, and its corresponding status is set to true. E.g., delete (abxd, abcd), and then after execution of delete (), input word becomes abd. Then, again the error is calculated, and the check is made if (error < maxError), then maxError = error; the number of steps is incremented by 1.

- (xi) insert(inputWords, availableWord) It stores word in data object, and the check is made for mismatch position. Length of the input word and available word is evaluated and stored in length1 and length2, respectively, and the loop is iterated for min (length1,length2) times. During each iteration, character at current index in both words is compared, and if they are same, then -1 is returned; else, the mismatch index is returned to insert () method for insertion at mismatch position, the new input word is framed, and its status is changed to true. E.g., insert(Abd, abcd), and then, insert will invoke and input word becomes abcd. The error is recalculated and the check is made if (error < maxError), then maxError = error; the number of steps is incremented by 1.
- (xii) replace(inputWords, availableWord) It stores word in data object, and the check is made for mismatch position. Length of the input word and available word is evaluated and stored in length1 and length2, respectively, and the loop is iterated for min (length1, length2) times. During each iteration, character at current index in both words is compared, and if they are same, then -1 is returned; else, the mismatch index is returned to replace () method to replace by available word character at that index, and the status is changed to true. The error is recalculated, and the check is made if (error < maxError), then maxError = error; the number of steps is incremented by 1. E.g., for (Abxy, abcd): delete(aby, abcd), insert(abcy, abcd), and replace(abcd, abcd) were invoked serially.
- (xiii) append(inputWords , availableWord) When append() is invoked, length of the input word and available word is evaluated and is stored in length1 and length2, respectively, substring (1,min(length-1)) is fetched, and the character is appended at minimum length index in dictionary word to input word. The data is set to the updated input word, and its status is set to true. The error is recalculated, and the check is made if (error < maxError), then maxError = error; the number of steps is incremented by 1.
- (xiv) trim(inputWords , availableWord) When trim () is invoked, it trims a character at a given position of input word with the respective character of available word, input word is set in data, and the status is set to true. The error is recalculated, and the check is made if (error < maxError), then maxError = error; the number of steps is incremented by 1.
- (xv) The steps ix–xiv are repeated till error=0.
- (xvi) At the end, the corrected word is used to reframe the user query which is used for further execution and retrieving the relevant information.

3.2 Algorithm of Spelling Correction

```

Input :
inputWord : city/state name in given query
Output:Available correct city/state name
replacementMap = Map to store new name of city against
old name of city.
dictionary : Array to store all cities and states.
procedure correctSpelling( inputWord )
1. inputWord = inputWord.lowerCase
2. replacementWord = get newWord against inputWord from
   replacementMap
3. if replacementWord != NULL then
4.     return replacementWord
5. end if
6. result = create empty String
7. minSteps = Max Integer value
8. for each String avblWord in dictionary do
9.     avblWord = avblWord.lowerCase
10.    steps = CALL SpelingCorrectionAlgo.apply( inputWord ,
   avblWord)
11.    totalError = CALL SpelingCorrectionAlgo.calculateError
   (inputWord,avblWord)
12.    if totalError == 0 then
13.        result = avblWord
14.        break
15.    else
16.        if steps < minSteps then
17.            minSteps = steps
18.            result = avblWord
19.        end if
20.    end if
21. end for
22. end procedure

```

4 Performance Analysis

The misspelled query terms are the main parameter that degrades the performances of any search interface. While entering the query in the semantic tourism search interface, if the user enters wrong spelling of a city or state, the novel spelling correction algorithm is implemented which successfully interpret the query by correcting the misspelled word. The sample comparative performance analysis with commonly

used search engines like Google, Bing, and Yahoo is discussed in detail in [15] which clearly illustrates how the use of spelling correction module remarkably improves the overall performance of the semantic search interface in terms of precision as and query execution time.

5 Conclusion

Spelling error is the main attribute that degrades the performances of any search interface. While entering the query in the semantic tourism search interface, if the user enters wrong city\state name, the system is unable to interpret the query and thereby does not provide any result. To resolve this, in this research, the performance of the semantic search interface is improved remarkably in terms of precision and execution time by implementing novel spelling correction module to handle misspelled cities\ states. Within the scope of this research, Indian Tourism is chosen as domain of study, but this spelling correction module can easily be implemented on any other domain or dictionary.

References

1. Lewandowski, D.: The retrieval effectiveness of search engines on navigational queries. Aslib Proc. **63**(4), 354–363 (2011)
2. Arnold, S.: Search: the new application platform. Electron. Lib. **24**(2), 121–125 (2006)
3. <http://searchenginewatch.com/article/2065173/How-Search-Engines-Work>
4. <http://en.wikipedia.org/wiki/Web-search-engine>
5. Laddha, S.S., Laddha, A.R., Jawandhiya, P.M.: New paradigm to keyword search: a survey. In: IEEE Xplore Digit. Libr. pp. 920–923 (2015). <https://doi.org/10.1109/ICGCIoT.2015.7380594>, IEEE Part Number: CFP15C35-USB, IEEE ISBN: 978-1-4673-7909-0
6. Vinayek, P.R., Bhatia, A., Malhotra, N.E.E.: Competitiveness of Indian tourism in global scenario. ACADEMICIA: An Int. Multidisciplinary Res. J. **3**(1), 168–179 (2013)
7. Hall, C.M.: Crisis events in tourism: subjects of crisis in tourism. Curr. Issues Tour. **13**(5), 401–417 (2010)
8. Laddha, S.S., Jawandhiya, P.M.: An exploratory study of keyword based search results. Indian J. Sci. Res. **14**(2), 39–45 (2017). ISSN: 2250-0138 (Online)
9. Laddha, S.S., Jawandhiya, P.M.: Onto semantic tourism information retrieval. Int. J. Eng. Technol. (UAE) **7**(4.7), 148–151 (2018). <https://doi.org/10.14419/ijet.v7i4.7.20532>. ISSN: 2227-524X
10. Nagata, R., Takamura, H., Neubig, G.: Adaptive spelling error correction models for learner English. Proc. Comput. Sci. **112**, 474–483 (2017)
11. Laddha, S.S., Jawandhiya, P.M.: Semantic search engine. Indian J. Sci. Technol. **10**(23), 01–06 (2017). <https://doi.org/10.17485/ijst/2017/v10i23/115568>
12. Laddha, S.S., Jawandhiya, P.M.: Semantic tourism information retrieval interface. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 694–697. <https://doi.org/10.1109/ICACCI.2017.8125922>
13. Laddha, S.S., Jawandhiya, P.M.: Novel concept of query prototype and query similarity for semantic search. In: Deshpande, A.V., Unal, A., Passi, K., Singh, D., Nayak, M., Patel, B., Pathan, S. (eds.) Smart Trends in Information Technology and Computer Communications.

- SmartCom 2017. Communications in Computer and Information Science, vol. 876, pp. 338–345. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-1423-0_35 ISBN: 978-981-13-1423-0 ISSN: 1865-0929
- 14. Laddha, S.S., Jawandhiya, P.M.: Novel concept of query similarity and meta-processor for semantic search. In: Bhatia, S.K., Tiwari, S.T., Mishra, K.K., Trivedi, M.C. (eds.) Advances in Computer Communication and Computational Sciences. IC4S 2017. Advances in Intelligent Systems and Computing, vol. 760, pp. 111–120. Springer, Singapore (2018). https://doi.org/10.1007/978-981-13-0344-9_9, ISBN: 978-981-13-0344-9, ISSN: 2194-5357
 - 15. Laddha, S.S., Koli, N.A., Jawandhiya, P.M.: Indian tourism information retrieval system: an onto-semantic approach. Proc. Comput. Sci. **132**, 1363–1374 (2018). ISSN 1877-0509. <https://doi.org/10.1016/j.procs.2018.05.051>

Cost-Effective NSX Inventory State Management in Cross-Cloud



Saloni Garg and Ravindra Kumar

Abstract With the extreme growth in technology, the amount of data that belong to an organization is many folded. It is practically impossible to persist entire data on premises. So many organizations are turning toward the cloud. In this paper, the focus is on cross-cloud environment where NSX manager which manages both on premise and public cloud workflow is deployed on premises of an enterprise customer. In such a situation, data must be sent to the public cloud environment. But there are certain fault scenarios which require whole data to be sent again and again to cloud. The aim is to reduce the amount of data being sent in such situations and find the current cost of maintaining inventory information.

Keywords Cross cloud · Full SYNC · Checksum · NSX manager · vSAN

1 Introduction

In traditional on premise deployments, as the scale of these host elements grows, it is impractical to persist all the attributes of the interested asset set on the NSX manager (centralized network management component). There are more practical ways to handle NSX crashes like to hold some of the data in memory and bring them back when NSX crashes. It will help to recover from NSX failure efficiently.

The NSX manager itself is deployed in a high availability configuration with shared storage, thereby reducing the risk of loss of data.

The product plans to support scenarios where these assumptions of repopulating the entire data set may not be cost-effective [1].

S. Garg (✉) · R. Kumar
Department of Computer Science and Engineering, Nalanda College of Engineering,
Chandi, Nalanda, Bihar, India
e-mail: gargsaloni31@gmail.com

R. Kumar
e-mail: ravi.sliet@gmail.com

Problem Statement

To find the current cost of maintaining inventory information for a typical NSX deployment, identify the fault scenarios which trigger repopulation of host-specific inventory information and propose a solution to effectively reduce the inventory delta exchange making efficient use of the expensive WAN bandwidth.

2 Ease of Use

2.1 Current Protocol (Existing System)

In cloud environment, whether it is on premise or public cloud, there need to send inventory data related to the objects that already exist. First time when data is being sent, it is fine to send the entire data set and after that only updates need to be sent to reduce the overhead on network.

DA (Discovery Agent) which sits on ESXi host will send SYNC_INIT_REQ to MP (management plane) through MPA (management plane agent) when there is a connection between MP and MPA. When MP is ready to receive full sync, it will send SYNC_INIT. At this point, entire host inventory information is sent to MP. Once host gets acknowledgment from MP, host got to know it has received what all data have sent. From now onwards, only updates will be sent (Fig. 1).

There are many full Sync scenarios (fault scenarios) for which there exists no way to find out what all data has been sent successfully. So, MPA always sends full sync (whole data).

Therefore, the aim is to identify full Sync scenarios (Fault Scenarios) which trigger repopulating entire information.

And the cases are challenges.

2.1.1 Hardware and Software Refresh

As the technology grows, there is a significant improvement in server's lifespan, but even with various options available for extending the server's lifespan, there is a need to purchase new hardware.

There comes a point, when server hardware is simply too old to perform, and again the question is when would you know when it is really time. To know about the new hardware purchase, the key is to identify that, when increased performance, energy efficiency requirements will justify a new hardware purchase [2].

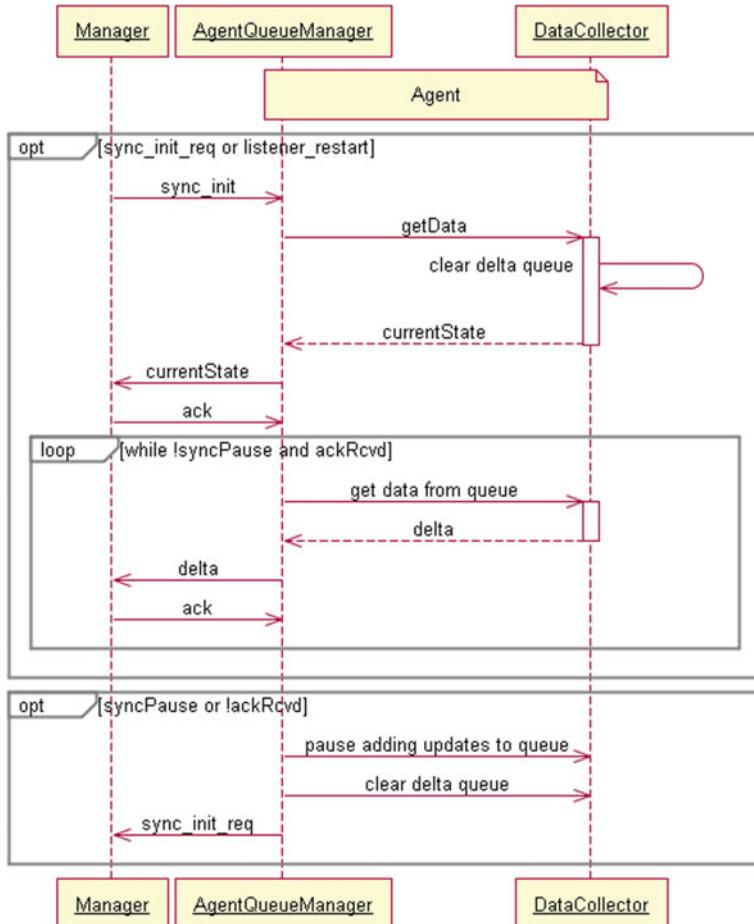


Fig. 1 Existing system view. Figure 1 is taken from <https://www.vmware.com>

2.1.2 VM Migration

Conversion from physical server to virtual server (VM) is called as physical to virtual migration.

For performing physical to virtual machine migration, we need to copy the physical server's bits to virtual disk, install drivers, and modify the guest to support these drives.

Workload keeps changing in servers. There is always a sudden spike of requests, i.e., once in a month. At that time, it is not possible to provision additional servers manually.

2.1.3 Unplanned Data center Downtime

- Human error and poor infrastructure capacity management.
- Poor maintenance and lifecycle strategy.
- Substandard data center site selection and risk mitigation.

2.1.4 Hardware/Server Failure

Servers which are presented in the data centers are always running and continuously heating up; this is the reason for h/w failure. There are some more reasons for equipment failure, like if temperature increases from its fixed value that decrease the server's performance and might be the reason for data center failure.

2.1.5 Power Outage

The reason for power outage is the uninterrupted power supply (UPS). It can happen from battery failure or from the excessive power draw which reaches beyond the predefined limits.

2.1.6 Host Goes Down

The reason for host breakage is the network traffic between the host servers and SAN (storage area network which is the location of VM file).

2.2 Proposed System

In the existing system, there need to send whole data again and again whenever any trash happens. So, there is a need to reduce this data transfer by storing sent data somewhere in the repository, and this is done by using checksum (Fig. 2).

DA (Discovery Agent) which sits on ESXi host will send SYNC_INIT_REQ to MP (management plane) through MPA (management plane agent) when there is a connection between MP and MPA. When MP is ready to receive full sync, it will send SYNC_INIT. At this point, entire host inventory information is sent to MP with the calculated checksum. Now onwards only checksums are being transferred rather than whole data. If checksum matches on manager side that means, there is no error in the sent data. If not, then manager's side calculated checksum is being transferred to the MPA for checksum comparison, and this process continues until whole data reaches to manager side safely. Now there is no need to send the whole file again and again.

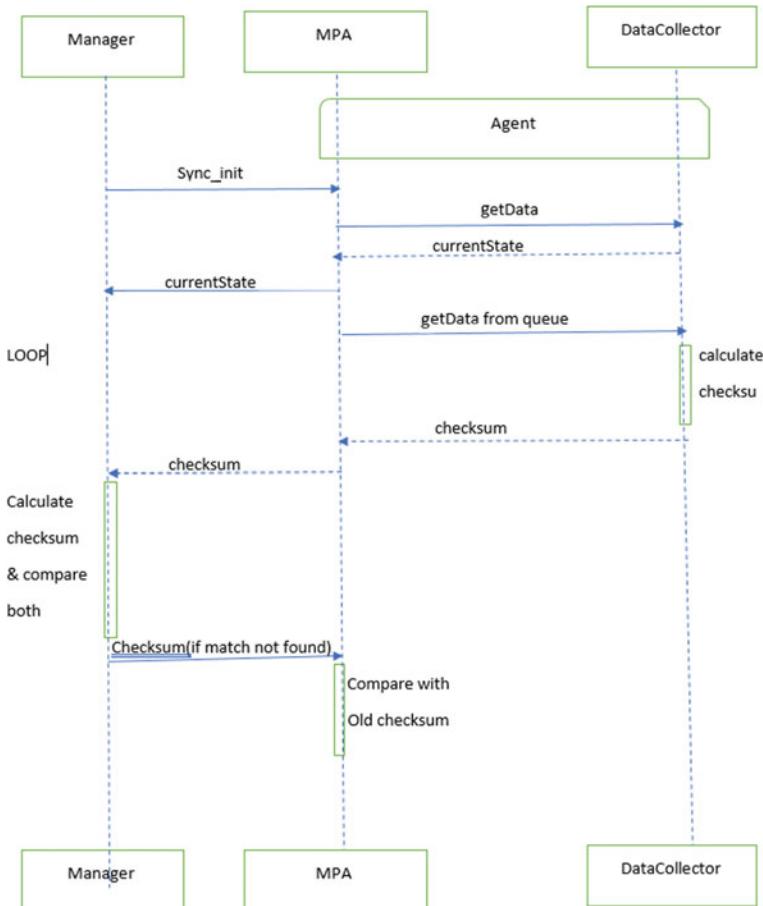


Fig. 2 Proposed system view. *Source Author*

3 Proposed Solution for Fault Cases

These are the solutions for the scenarios where current protocol fails and needs to send full data again and again to the cloud [3].

3.1 *Hardware/Server Refresh*

In some cases, a server hardware update can solve performance issues at the fraction of the price of a new server. Adding CPUs or memory can significantly increase server

performance. However, we cannot fix all the hardware by upgradation because not all the hardware is upgradable. There is always a need for new purchase.

3.2 VM Migration

Virtual machine migration accommodates for changing workloads automatically. It can manage to both additional workload and reduced workload.

For the users of servers, a scheduled maintenance normally results in some downtime. Virtual migration helps to migrate virtual machines from one host to another and brought back to the original server when migration is completed.

If there is a case of unscheduled server downtime, virtual machine migration can also migrate from one host to another host. (This is the case of any random server fault).

For disaster recovery, virtual machine migration can also be used. This process involves setting up of similar resources in a disaster recovery site with high-speed wide area network links and specializes network connectivity equipment. This process involves setting up of similar resources in a disaster recovery site with high speed Wide Area. Network links and specializes network connectivity equipments between primary server and the servers in the data recovery site, the virtual machines and their array.

Comparative to migration of operating systems and applications, migration of virtual machine is a quite effortless process.

With the help of virtual machine migration, it is possible to migrate operating system from older server to newer one easily and without disrupting the services [4].

3.3 Unplanned Data center Downtime

This can be managed by buying better quality hardware from the known vendors.

3.4 Power Outage

In the power outage, the UPS system draws from its battery backups, making them an essential piece to maintain uptime. However, batteries do not last forever. Batteries should be tested at least quarterly for verification of correctness.

Cyberattacks may be a reason for power outage. CRAC failure has also become increasingly common as densities rise.

3.5 Host Goes Down

The suggested solution for host failure is VMware high availability (HA). It helps by gathering all the VMs and hosts on which VMs present. HA first checks for failed host, and then, it can restart a VM which is running on a failed host. HA can also be checked by checking the VMware tools which are running on it. When any ESX/ESXi host (Host on which VMs are running) fails for any reason, then VMs can also fail. VMware HA ensures that VMs which are running on the failed host can run on another host (means they should be capable of being restarted).

ESX/ESXi host failure is primarily dealt with the VMware high availability and what happens with the virtual machines that are running on the host. HA does not mean fault tolerance. These are two different terms, HA checks availability for VMs which failed because of host failed and fault tolerance focuses on making system more fault tolerance.

4 Size Calculation

This is the size of the message which is being sent between MPA and MP with the 10,000 VMs and 512 hosts. The number of VMs and Hosts can vary (Table 1).

The total size of the message is 17,958 KB (Approx.).

When data from multiple hosts is coming, the cloud receives a large amount of data. So, it is very inconvenient to repopulate this entire data in case of any failure.

Procedure

A host comprises of multiple virtual machines. The virtual machines further have multiple virtual network interfaces. So in order to calculate size of one host there is need to add the size of host to the size of all the virtual machines. Size of one virtual machine is already known so calculate the size of actual number of virtual machines that present on one host. To calculate the size of virtual machines, there is a need to include the size of virtual network interfaces that is available per virtual machines. By doing all the calculations using unitary method, the size of one host is computed. Finally, we can calculate the size of multiple hosts. This size will in turn help to find the overall size of the message that is being sent from management plane agent to management plane [5].

Table 1 Size of messages for proposed system

	Number	Size/object (in KB)
VM	10,000	0.24
VIF	50,000	0.38
Host	512	35

Source Author

Size of IDA message per host

$$\begin{aligned} \text{Size} = & (((\text{no of VIFs} * \text{size of VIFs}) \\ & + \text{size of VM} * \text{no of VMs}) \\ & + \text{size of Host} * \text{no of hosts}) / (1024) \end{aligned}$$

Size of IDA message for total no of hosts

$$\begin{aligned} \text{Size} = & (((\text{no of VIFs} * \text{size of VIFs}) \\ & + \text{size of VM} * \text{no of VMs}) \\ & + \text{size of Host} * \text{no of hosts}) / (1024) \end{aligned}$$

The calculated size of IDA message for one host is 35 KB
 And the size of IDA message for multiple hosts or one cluster is 17,958 KB (Approx.).

5 Conclusion and Future Work

In this paper, Most of the fault scenarios are being discussed. So that after analyzing these scenarios, repopulation of host inventory can be minimized and there is less need to pay high expenses of sending large amount of data over cloud.

These fault scenarios are based upon some cases where full SYNC is being sent from MPA to MP using data collector and data sender.

Now need to find an optimal method to reduce the amount of data to be sent across cloud in these scenarios. For that calculation of checksum and version modeling will be done for all the updates. So that tracking of sent data and remaining data will be performed. On comparing it with the latest file checksum value, will get to know that how much data needs to be sent.

References

1. Anil, S.L., Thanka, R.: A survey on security of data outsourcing in cloud. *Int. J. Sci. Res. Publ.* **3** (2013)
2. Fildes, R., Beard, C.: Forecasting systems for production and inventory control. *Int. J. Oper. Prod. Manag.* **12**(5), 4–27 (1992)
3. Theng, D., Hande, K.N.: VM management for cross-cloud computing environment. In: 2012 International Conference on Communication Systems and Network Technologies (CSNT), pp. 731, 735, 11–13 May 2012

4. Li, Y., Xu, X., Ye, F.: Research on supply chain inventory optimization and benefit coordination with controllable lead time. In: Wireless Communications, Networking and Mobile Computing, 2007
5. Selvarani, S., Sudha Sadhasivam, G.: Improved cost-based algorithm for task scheduling in cloud computing. In: International Conference on IEEE, Computational Intelligence and Computing Research, 2010

A Tri-band Staircase Antenna for RF Energy Harvesting



Aysha Munir Sheikh and Rajeev Mathur

Abstract A staircase-shaped antenna is presented in this paper for the purpose of RF energy harvesting. The antenna is designed and simulated, and the results are inquired with reference to return loss, VSWR, gain, and polar plot. The proposed antenna is fabricated on a low-cost FR4 substrate, and measurements are taken using VNA. This antenna has a length of 28.30 mm (L) and width of 38.03 mm (W) and is excited with a microstrip line feed technique. The antenna is a tri-band antenna, operating at the resonant frequencies of 2.53, 4.46, and 5.5 GHz. These frequencies are used for Bluetooth, WLAN, and WiMAX wireless communication. The simulated results and measured results are in satisfactory concordance with each other. This antenna, if used with DC rectifier circuit, could be a good candidate for the use of RF energy harvesting.

Keywords RF energy harvesting · Rectenna · DC rectifier · Return loss · Staircase antenna · Bow-tie antenna

1 Introduction

Due to extensive use of wireless communication nowadays, ample RF energy is freely available in the environment. This energy could be captured and be used for low power devices like mobile phones and Wi-Fi dongles. Energy Harvesting is a method of grasping the accessible energy from the exterior circulatory sources. Several research groups are going to harvest energy from these RF signals. This technique fascinated great emphasis and receiving antennas, matching circuits, and rectifying circuits carried out by various RF energy harvesting systems which have

A. M. Sheikh (✉) · R. Mathur
Geetanjali Institute of Technical Studies, Udaipur, Rajasthan, India
e-mail: ayshamunir12@gmail.com

R. Mathur
e-mail: mathur.rajeev96@gmail.com

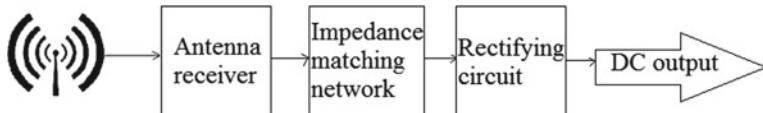


Fig. 1 Block diagram of RF harvesting system [3]

been flourished for green supply of low power electronics. This technology is grown to have battery-less wireless devices [1].

For RF harvesting, antenna is the basic device used to capture radio wave energy by using antenna. This RF energy has hugely developed because of the great part of wireless signals. Rectenna is also known as a rectifying antenna which is used to transform electromagnetic energy into direct current electricity [2].

Figure 1 shows that in this system the network controller or ambient source transmit the RF energy which is elevated by the antenna, the impedance matching network provides maximize the power transfer or minimize signal reflection from the load, and the rectifier transforms the RF signal to direct current (DC) signal, which flows in only one direction [3].

Thus, three major components are needed for the RF harvesting system, i.e. an antenna receiver, matching network, and rectifying circuit. It is required to properly design these units to maximize harvesting ambient RF energy; for this, we need to have the following:

- i. A well-organized antenna to capture radio frequency energy.
- ii. A matching network exists to match the characteristic impedance or antenna impedance to prevent standing waves and to provide efficient transfer of power from source to load.
- iii. A rectifier circuit using special diodes, for RF to DC conversion [3].

2 Antenna Design

Dimensions of the proposed antenna are calculated through standard designing method. By using this method, the size of the radiating patch is determined as $28.30 \text{ mm} \times 38.03 \text{ mm}$ and it is fed by a microstrip line. Length and width of staircase cut are $5.66 \text{ mm} \times 4.2255 \text{ mm}$. Parameters for designing antenna are as follows: thickness of substrate, 1.6 mm; relative dielectric constant, 4.4; and resonant frequency, $f_r = 2.4 \text{ GHz}$. Figure 2 shows the structure of the designed antenna.

The design specifications for the energy harvesting novel proposed tri-band antenna and a ground plane structure at 2.4 GHz resonant frequency are as follows:

- Width of designed antenna (W) = 38.03 mm
- Length of designed antenna (L) = 28.30 mm
- Width of ground plane (W_g) = 47.63 mm

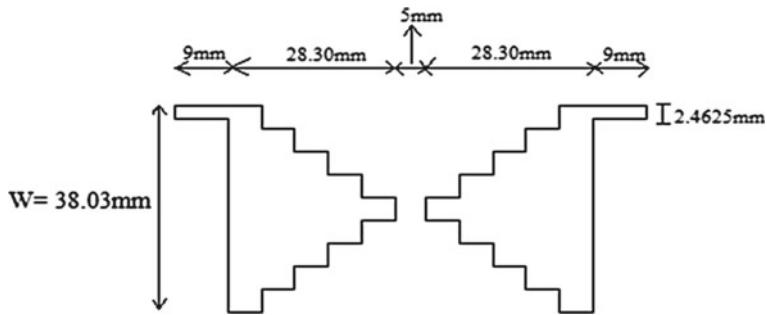


Fig. 2 Structure of staircase antenna

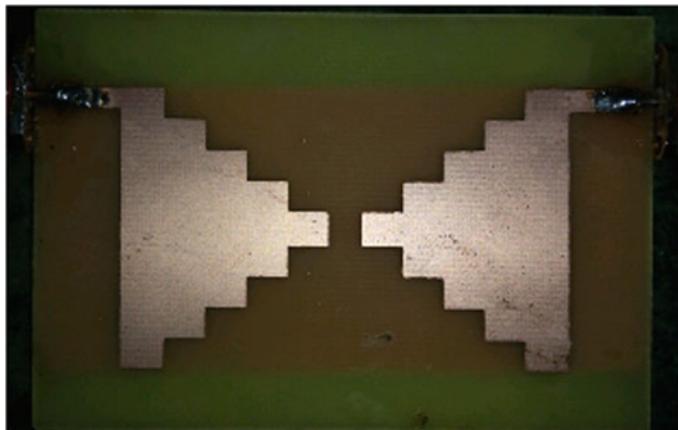


Fig. 3 Fabricated staircase antenna for RF energy harvesting

- Length of ground plane (L_g) = 79.6 mm
- Dielectric constant of FR4 substrate (ϵ_r) = 4.4
- Height of substrate (h) = 1.6 mm
- Loss tangent = 0.00026
- Feed line length (L_f) = 9 mm
- Feed line width (W_f) = 2.4625 mm
- Overall dimensions of the rectenna are width (W) = 38.03 mm and Length (L) = 56.60 mm.

The antenna is constructed on easily available low-cost FR4 substrate and the fabricated staircase bow-tie antenna as shown in Fig. 3.

3 Simulated and Experimental Results

In this segment, simulated and experimental results of the proposed antenna are discussed. The antenna is simulated using commercially available software, and simulated results are acquired. The constructed antenna is tested on Anritsu VNA, and the results are compared with the simulated results.

3.1 Return Loss

Simulated return loss indicates a response at 2.47 GHz is -24.78 dB, 4.39 GHz is -17.04 dB, and 6.49 GHz is -12.50 dB, respectively, as displayed in Fig. 4. Figure 5 presents the measured result of Return Loss by using VNA.

3.2 VSWR

VSWR plot for staircase antenna is shown in Fig. 6, which determines the matching properties of antenna. Voltage standing wave ratio (VSWR) lies in the range of 0.9–1.49 for the frequencies of 2.47, 4.39, and 6.49 GHz, i.e., 0.9 for 2.47 GHz, 0.9 for 4.39 GHz, 1.49 for 6.49 GHz. Figure 7 shows the measured result of VSWR by using VNA.

3.3 Radiation Pattern

Generally, radiation pattern can be defined as the function of angular position and radial distribution from the antenna or as the power transmitted or received by the antenna. The simulated radiated field pattern at the center frequency $f_c = 2.4$ GHz for azimuth and elevation plane is shown in Fig. 8. The peak gain is obtained around

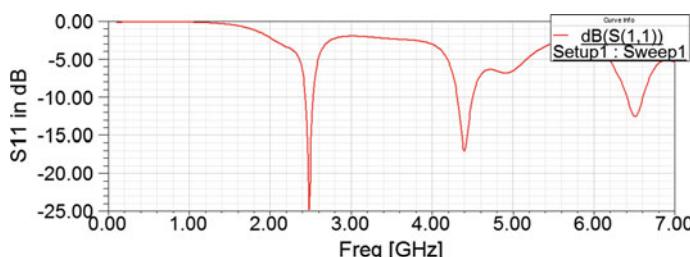


Fig. 4 Simulated return loss of staircase antenna

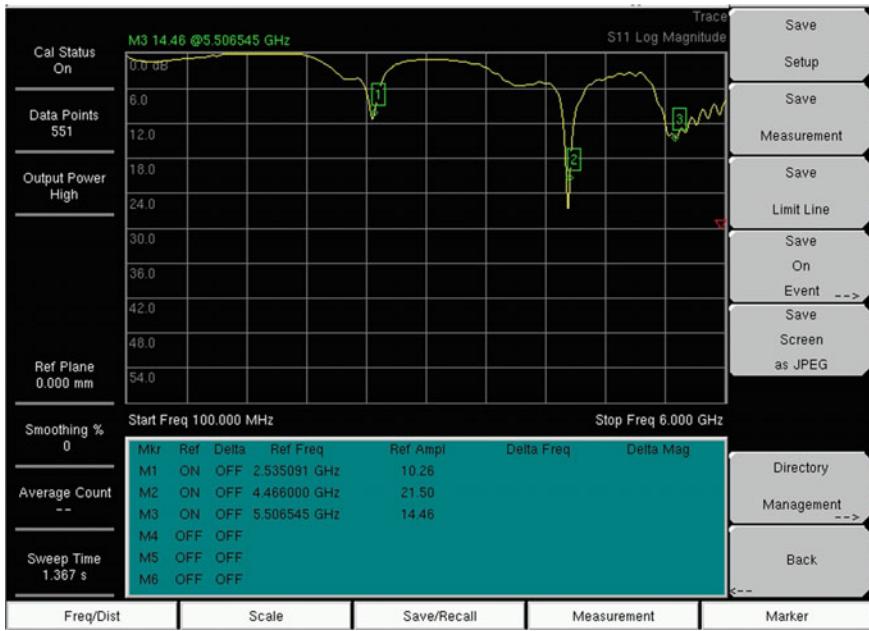


Fig. 5 Return loss of staircase antenna measured on VNA

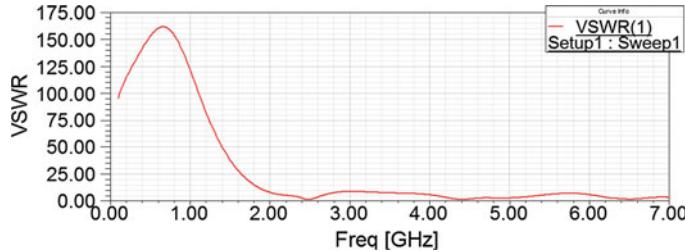


Fig. 6 Simulated VSWR of staircase antenna

11.27 dB at this frequency. 3D Polar Plot of radiation pattern at $F_c = 2.4$ GHz is shown in Fig. 9.

3.4 Comparison Between Simulated and Experimental Results

Table 1 depicts that there is a satisfactory agreement between simulated and experimental results

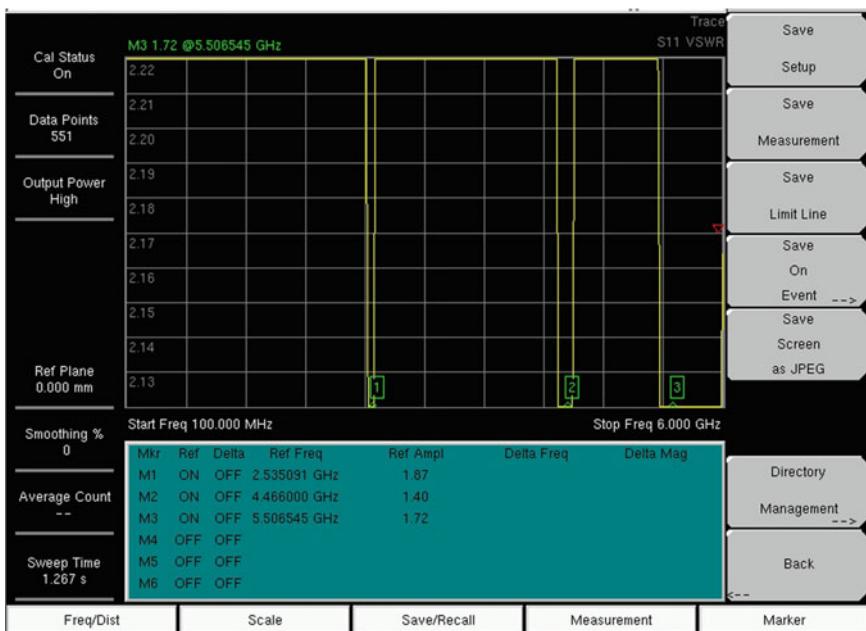


Fig. 7 VSWR of staircase antenna measured on VNA

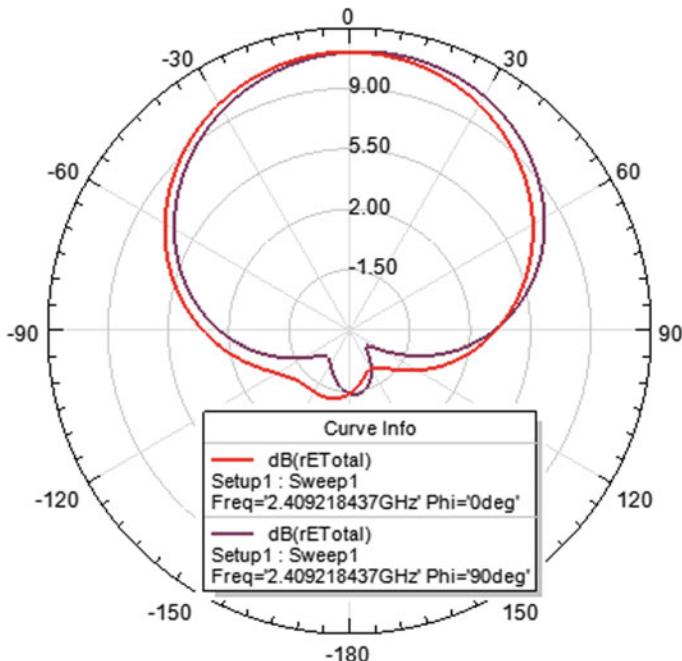


Fig. 8 Simulated radiation field pattern of staircase antenna at 2.4 GHz

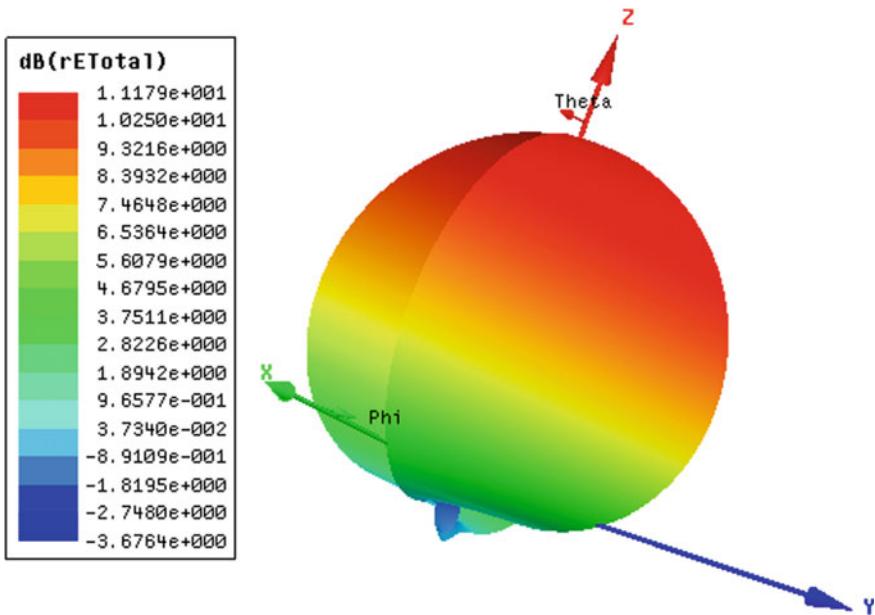


Fig. 9 3D radiation pattern at 2.4 GHz

Table 1 Comparison between simulated results and experimental results

Parameters	Simulated results	Experimental results
Resonant frequency	2.47 GHz	2.53 GHz
	4.39 GHz	4.46 GHz
	6.49 GHz	5.50 GHz
Return loss	24.78 at 2.47 GHz	10.26 at 2.53 GHz
	17.04 at 4.39 GHz	21.50 at 4.46 GHz
	12.50 at 6.49 GHz	14.46 at 5.50 GHz
VSWR	1.78 at 2.47 GHz	1.87 at 2.53 GHz
	1.24 at 4.39 GHz	1.40 at 4.46 GHz
	1.50 at 6.49 GHz	1.72 at 5.50 GHz
Gain	11.27 dBi	5.96 dBi

4 Conclusion

One of the most important components of rectenna is antenna, and the changes on its design can harvest more RF energy from circulating sources. We have successfully designed, fabricated, and tested a novel tri-band antenna with center frequency of 2.4 GHz and 4.4 GHz and 5.5 GHz for RF harvesting. Here staircase-shaped antenna is used for significant reduction in antenna size. Return loss at these frequencies is below -10 dB, and the VSWR of the proposed antenna is also within the range of 1–2 in above frequencies. We have considered matching circuit requirements by studying

Smith chart, and this antenna has impedance nearly 50 ohms fulfilling the required impedance. The main advantage of the proposed antenna is that it can capture RF energy from three ambient frequencies. Gain of the antenna is also significantly high thereby making it more energy efficient.

5 Future Work

A matching network could be added with the proposed antenna to maximize power transfer between the antenna and rectifier circuit. DC output of rectifier circuit could be obtained using a good quality Schottky diode, and efficiency of the RF harvesting system could be evaluated. This DC voltage can be used for low power applications and make RF energy harvesting a good solution for battery-less applications.

References

1. Viji, P., Nagasaratha, P., Sriranjini, R.: Energy harvesting using slot antenna at 2.4 GHz. *Int. J. Antennas (IJANT)* **2**(2) (2016)
2. Patil, S., Gahankari, S.: Design and implementation of microstrip antenna for RF energy harvesting. *Int. J. Eng. Res. Technol.* **10** (2017)
3. Bakkali, A., Pelegri-Sebastia, J., Sogorb, T., Llario, V., Bou-Escriva, A.: A dual-band antenna for RF energy harvesting systems in wireless sensor networks. *Hindawi Publ. Corp. J. Sens.* (2016)
4. Parekh, S., Mathur, R., Jain, P.: Triband handset antenna designing with a high gain for UMTS/WiMAX/WLAN applications. In: Vishwakarma, H., Akashe, S. (eds.) *Computing and Network Sustainability. Lecture Notes in Networks and Systems*, vol 12. Springer, Singapore (2017)
5. Leclerc, C., Egels, M., Bergeret, E.: Design and measurement of multi-frequency antennas for RF energy harvesting tags. *Prog. Electromagn. Res.* **156** (2016)
6. Shrestha, S., Noh, S.-K., Choi, D.-Y.: Comparative study of antenna designs for RF energy harvesting. *Hindawi Publ. Corp. Int. J. Antennas Propag.* (2013)
7. Sreedhar Kumar, G., Praveen Kumar, P.C., Sarath Manohar Babu, B.A., Naresh Babu, B: Dual staircase shaped microstrip patch antenna. *Int. J. Adv. Res. Comput. Commun. Eng.* **3**(6) (2014)
8. Berges, R., Fadel, L., Oyhenart, L., Vigneras, V., Taris, T.: A dual band 915 MHz/2.44 GHz RF energy harvester. In: 45th European Microwave Conference, Paris, France, pp. 307–310, Sept 2015, HAL Id: hal-01475276. <https://hal.archives-ouvertes.fr/hal-01475276>
9. Patel, D., Mehta, R., Patwa, R., Thapar, S., Chopra, S.: RF energy harvesting. *Int. J. Eng. Trends Technol.* **16** (2014)
10. HFSS by Ansys: <http://ansys.com/en-GB/Products/Electronics/ANSYS-HFSS>
11. Kim, S.: RF energy harvesting techniques for wirelessly powered devices. In: Proceedings of the IEEE MTT-S International Microwave Workshop Series on Intelligent Radio for Future Personal Terminals, Daejeon, Republic of Korea, 2011
12. Din, N.M., Chakrabarty, C.K., Bin Ismail, A.: Design of RF energy harvesting system for energizing low power devices. *Prog. Electromagn. Res.* **132** (2012)
13. Pinuela, M., Mitcheson P.D., Lucyszyn, S.: Ambient RF energy harvesting in urban and semi-urban environments. *IEEE Trans. Microw. Theory Techn.* **61**(7), (2013). <https://doi.org/10.1109/TMTT.2013.2262687>. <https://www.researchgate.net/publication/260333253>

14. Mathur, R., Joshi, S.: A split ring array antenna for WLAN applications using MIMO techniques, special issue on evolution in network and computer communications. *Int. J. Comput. Appl.* 30–44 (2011)
15. Shukla, K., Makwana, R.: A survey on different rectenna designs as an energy harvester. *Int. J. Innov. Res. Comput. Commun. Eng. (IJIRCCE)* 4(2), (2016). ISSN (Print): 2320-9798, ISSN (Online): 2320-9801
16. Mathur, R., Joshi, S.: A novel multiple element patch antenna for wireless MIMO beamforming & WiMAX applications. In: International Conference on Emerging Trends in Networks and Computer Communications (ETNCC), pp. 110–114 (2011)
17. Ahmed, S., Husain, M.N., Zakaria, Z., Zin, M.S.I.M., Alhegazi, A.: Rectenna designs for RF energy harvesting system: a review. *Int. J. Commun. Antenna Propag. (I.Re.C.A.P.)* (2016). <https://doi.org/10.15866/irecap.v6i2.8736>, <https://www.researchgate.net/publication/303845511>
18. Shao, S., Gudan, K., Hull, J.J.: A mechanically beam-steered phased array antenna for power-harvesting applications [antenna applications corner]. *IEEE Antennas Propag. Mag.* **58**, 58–64 (2016)
19. Lakhal, H., Dhib, M., Ghariani, H., Lahiani, M.: An efficient rectenna for RF energy harvesting applications at 2.45 GHz. *Int. J. Sci. Tech. Automatic Control Comput. Eng. (IJ-STA)* **10**(Special Issue ESA), 2109–2113 (2016)
20. Balanis, C.A.: *Antenna Theory: Analysis and Design*. 3rd ed., pp. 811–865. Wiley-Interscience, Hoboken, NJ (2005)
21. Tran, L.-G., Cha H.-K., Park, W.-T.: RF power harvesting: a review on designing methodologies and applications. *Open Access, Micro and Nano Syst. Lett.* <https://doi.org/10.1186/s40486-017-0051-0>
22. Elsheakh, D.M.N.: *Microwave Antennas for Energy Harvesting Applications*. INTECH (2017). <http://dx.doi.org/10.5772/64918>

Aspect-Based Sentiment Analysis Using Deep Learning Convolutional Neural Network



Ravindra Kumar and Saloni Garg

Abstract Sentiment analysis (known as opinion mining) is the computational study of unstructured textual information regarding a person's perspective, attitudes, feeling, and emotions toward an event or an entity in the form of a piece of text. Sentiment classifiers automatically classify text in a positive and negative sense. In this process, two important aspects are presented. First is the feature extraction; it provides effective information about domain; and in our case, we use supervised learning with labels. The second aspect is training and validation in which tenfold cross-validation is done by using training and testing module with deep learning convolutional neural network. In this thesis, semantic feature is extracted by ontology and these features are trained in the convolution neural network, after that, tenfold cross-validation is done. In our experiments, CNN shows a significant difference in 72.17%, 71.66, and 71.66 in accuracy, precision, and recall, respectively.

Keywords Opinion mining · Sentiment · Machine learning · Deep learning

1 Introduction

Sentiment analysis is always a challenging task in the field of natural language processing. It is the process of finding opinion attached with a piece of text by evaluating whether it is positive, negative, or neutral. It also finds the degree of polarity (high, mild, or moderate) well known as opinion mining. It analyzes the feeling, thought, and attitude from collecting opinion as review on various web sites. Opinion mining can be done mainly by two approaches: first one is a machine learning approach, and the second one is a lexicon-based approach. Further machine learning-based approach is divided into two parts: supervised and unsupervised. There are three ways to perform sentiment analysis. First one is document level. It does not consider the diverse points of the archive; however, it can have moderately great outcomes particularly in cases

R. Kumar (✉) · S. Garg

Department of Computer Science and Engineering, NCE Chandi, Nalanda 803101, Bihar, India
e-mail: ravi.sliit@gmail.com

S. Garg

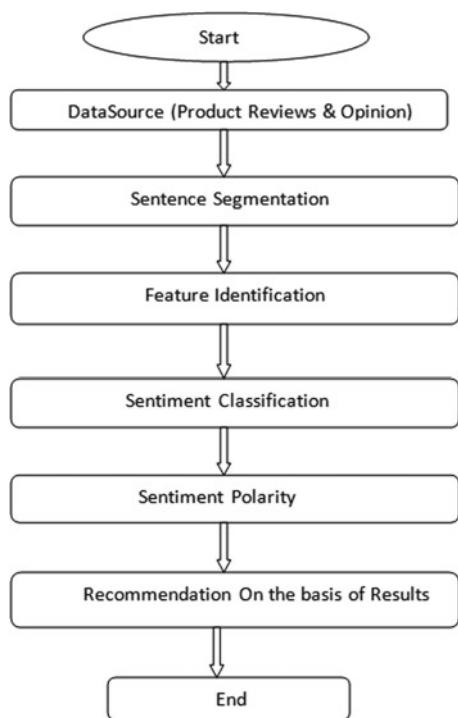
e-mail: gargsaloni31@gmail.com

where we are intrigued just in discovering which records do a particular client need to peruse. At sentence level, the sentence-level grouping considers each sentence as a different unit and expects that sentence ought to contain just a single supposition. Sentence-level examination has two undertakings as subjectivity characterization and estimation grouping. In phrase level, also known as aspect level, the objective of performing classification at feature level is to create a component-based perspective outline of different surveys. It has basically three undertakings. The primary assignment is to recognize and extricate questions included that have been remarked on by a conclusion holder (e.g. “picture,” “battery life”). The second job is to decide the extremity of viewpoints on different classes of features: positive, negative, and impartial. The third job is identified with the gathering words equivalent to features.

2 Related Work

Continuous effort has been taken to find out the exact polarity for reviews, but it is remained challenging to get better results from handling all unwanted conditions. Reported a series of experiments with convolutional neural network (CNN) trained

Fig. 1 Control flow of sentiment analysis



on top of pretrained word vector for sentence-level classification tasks. In this, a simple CNN with little hyper-parameter tuning and static vectors achieves excellent results on multiple benchmarks. It is implemented with single convolutional layer architecture. It can reduce the manual effort, but there may be probability in little loss in accuracy [1]. In this work, pre-trained word vector embedding is obtained by implementing unsupervised learning on large corpora. The data set used belongs to SemEval 2015, result supports the Twitter SemEval 2015 benchmark, and no hand-crafted features were utilized in this framework. The measuring F1 score was 64.85% [2]. In this architecture, three pairs of convolutional layer and max-pooling layer were used. This was the first time that seven-layer framework model was applied using word2vec and CNN to find out the sentiment of the sentence. Parametric Rectifier Linear Unit with normalization and dropout technology. Publicly available movie review data was used with five different labels: negatively, somewhat negative, neutral, somewhat positive, and positive. The network had acquired a test accuracy of 45.4% [3]. The neural network model uses a convolutional neural network with tree bank information for performing sentiment analysis task. This model uses a number of aspects like syntax information, structure information that works better than the other single aspect model [4]. A general approach is taken to do opinion mining at both the sentence level and complete review level. Proposed promises to perform better accuracy and significant [5]. A framework in which sentiment polarity detection has been done by CNN is proposed. Multiple implementations of the convolutional neural network are done with different filter size and different no of filter size, and then, AdaBoost is used to combine the different implementation of CNN to find the polarity of sentiment in the given sentence. Datasets used in this work are IMDB, and Rotten Tomatoes movie reviews already divided into two classes: positive and negative [6]. A model which studies Aspect based opinion summarization (AOS) of reviews on a specific product. This approach maps each sentence with predefined aspect. The discussed model consists of two CNN-based methods: cascaded and multitask. Level 1 deals with aspect mapping and a single CNN; level 2 deals with sentiment classification. For training purpose, the dataset used is Amazon smartphone reviews [7]. Multiple kernel learning architecture is implemented to combine audio, visual, and textual modularities. The proposed framework achieved accuracy in multimodal sentiment analysis with a margin of 10–13 and 3–5% accuracy in polarity detection and emotion identification. The future scope of this framework is to focus more on extracting relevant feature from visual content. This can be implemented by using 3D CNN implementation for automatic feature extraction [8]. Sentiment analysis was done by using probabilistic classifier in which naive classifier is used for finding the polarity attached to the review. It performs very well good classification results obtained with a tree-based feature selection and applying hierarchical cluster measures [9]. Detection of human emotion variation in an adolescent age class is done with the help of data mining technique. Different age group shows different emotions from their face expression [10]. Sentiment analysis of online feedback is done with the help of maximum entropy, and the SVM classifier outperforms other existing models [11]. Designed a system an image, the sentiment identification framework was built with a deep convolutional neural network. This framework was retrained

on large data for object recognition and transfer learning. The dataset used Flickr labeled images [12]. The whole sentiment analysis is represented in Fig. 1.

3 System Model

Each and every aspect of the proposed framework for sentiment analysis is represented in Fig. 2.

4 Problem and Solution Approach

4.1 Problem Statement

Sentiment analysis rapidly evolving research area, especially since social network comes to marketing. Sentiment analysis plays an important role in the analysis of user opinion for industry product, movie review, and understanding the behavior of user for specific products, but analysis of accuracy plays an important role in automatic analysis of behavior of the users. The present work focused on the ontology base feature extraction which adds the semantic information of the domain and then makes classifier model by using a convolution neural network with training and ten cross-validation. Semantic features and multilayer convolution neural network improve the training and testing accuracy of the complete model, and the result is far better than other models.

4.2 Solution Approach

4.2.1 Data Collection

For various three reasons, a test domain for a movie review is selected. Firstly, movie review has been utilized in previous ontology research frequently. Publicly available movie review provides higher quality reviews (such as Twitter). For example, IMDB archives are extracted for review newsgroups, the balanced negative and positive review comprises in MovieLens datasets of higher quality, the manual validation result of the polarities of sentiment and incomplete and non-English review cleaning manually. Secondly, common domain is movie, and ontology based on movie review accordingly is easily understood by the general public. Thirdly, some movie review mining-associated special challenges present an opportunity to deepen our problem understanding and newer solutions proposal.

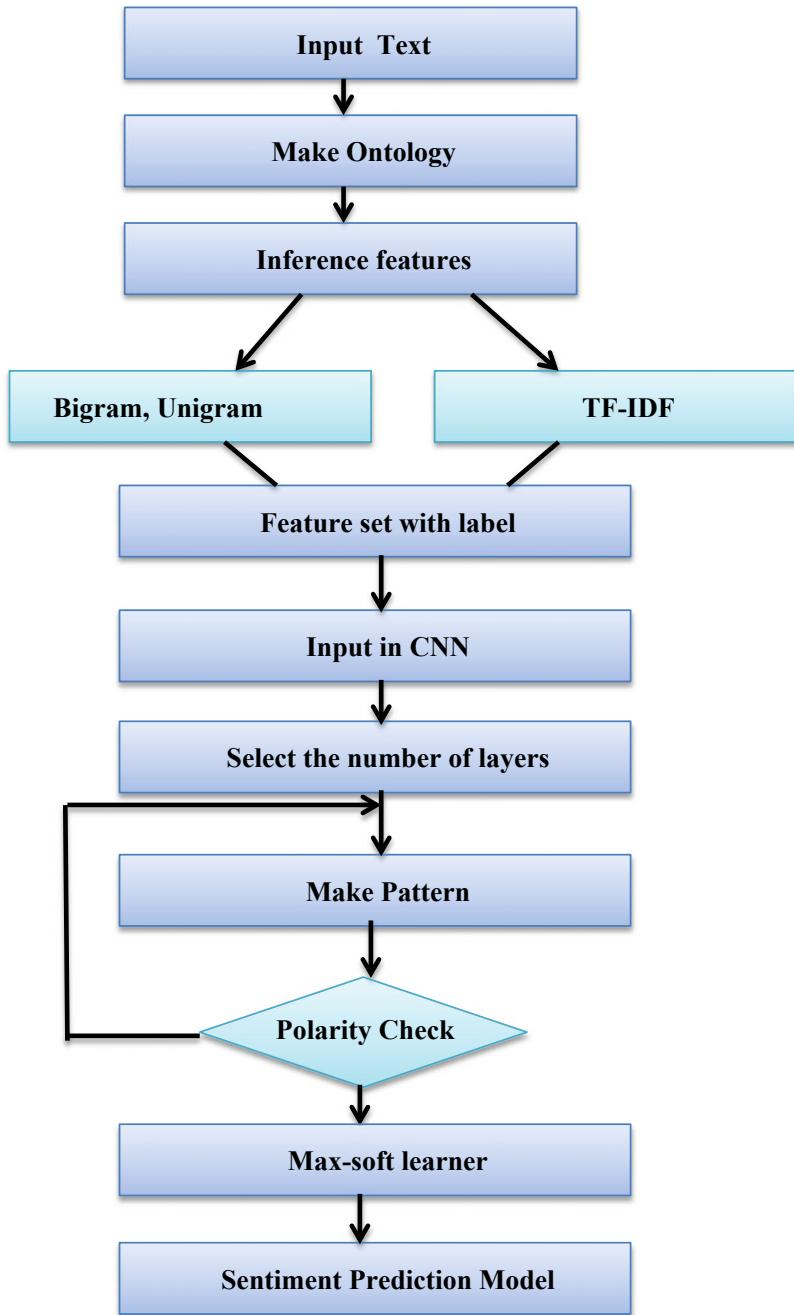


Fig. 2 Phases of the aspect-based sentiment analysis using deep learning CNN

4.2.2 Data Preparation

The movie reviews are reused which were analyzed manually in our experiment during the development of ontology, and the dataset that comprises movie review is grouped into negative and positive categories. In two versions, the dataset is prepared: segmented and original. The original reviews entirely are taken under original version as units that are treated as baseline. Ontology support enabled the review segments consisted in the segmented version that are generated on the basis of the movie review ontology properties.

4.3 Aspect-Based Ontology Framework

Ontology has a large history referring to the existing subject. In knowledge management context, an ontology is defined as some domains shared understanding that conceives often entities set, functions, instances, relations, and axioms. Ontology for knowledge reuse in an issue technologically is essential. Nowadays, requirement in the fields of tools development, integration, standardization, and users' adoption is high. For the requirement satisfaction, software applications have been developed which utilize Web Technology. A web-based system can avail some audit outcomes.

4.4 Deep Learning Convolutional Neural Network

For a given sentence, SCNNchar calculates each sentiment label score $\tau \in P$. For scoring a sentence, the word sequence in the sentence is taken as input in a network and it passes through the layer sequence in which the features are extracted with an increased complexity level. The feature extraction of network can be done from the sentence level and character level. The main network architecture novelty is two convolutional layers inclusion that allows it to handle any size sentence and words.

4.4.1 Initial Level Representation

The network's first layer converts words into feature vectors of real value (embedded) which captures syntactic, semantic, and morphological information about words. The word vocabulary of fixed size is v^{word} , and considering the words comprising of characters from character vocabulary of fixed size is v^{char} . The sentence comprises of n words $\{W_1, W_2, \dots, W_n\}$, and converting the W_n (each word) into vector $V_n = [r^{\text{word}}, r^{\text{wchar}}]$ is done using two sub-vectors: the embedding word level $r^{\text{word}} \in R^{d^{\text{word}}}$ and the embedding character level $r^{\text{wchar}} \in R^{d_n^0}$. The embedding of word level is meant for capturing semantic and syntactic information, and embedding of character level is meant for capturing the shape and morphological information.

4.4.2 Character-Level Embedding

Column vectors encode the word-level embedding in an embedding matrix $w^{\text{word}} \in R^{d^{\text{word}}} \times |v^{\text{word}}|$. The embedding of word level corresponds to each column $w_i^{\text{word}} \in R^{d^{\text{word}}}$ of i th vocabulary word. A word W transforms into embedding of word level by utilizing the product of matrix-vector:

$$r^{\text{word}} = v^w w^{\text{word}}$$

where

$v^w \leftarrow$ vector size of $|v^{\text{word}}|$ having 1 value at index w and 0 in other positions

The w^{word} matrix is a parameter for learning, and the embedding word-level size d^{word} is the user choosing the hyper-parameter.

4.4.3 Word-Level Embedding

The robust methods for extracting shape and morphological information from words taken into all characters consideration of the word and selecting the important feature for task. For instance in the sentimental analysis task of Twitter data, and appearing important information in different hashtag parts (e.g., “#ilikeit”) and various adverb information endings having suffix “ly” (e.g., “badly”). The local features are produced by convolutional approach around word’s each character, and further max operation is utilized for combining them for creating embedding character-level of fixed-size word.

Given words composed of m character $\{k_1, k_2, \dots, k_m\}$, each character k_m is transformed into embedding character r_m^{char} . Column vectors encode the embedding character in the matrix embedding $w^{\text{char}} \in R^{d^{\text{char}}} \times |v^{\text{char}}|$. Given k character, the matrix-vector product obtains its embedding r^{char} :

$$r^{\text{char}} = v^k w^{\text{char}}$$

where

$v^k \leftarrow$ vector size of $|v^{\text{char}}|$ having 1 value at k index and 0 in other positions.

The convolution layer input is the embedding character sequence $\{r_1^{\text{char}}, r_2^{\text{char}}, \dots, r_n^{\text{char}}\}$. The matrix-vector operation is applied by convolution layer for each window having size k^{char} of the sequence of successive windows. Consider the defined vector as $x^m \in R^{d^{\text{char}}} R^{d^{\text{char}}}$ which is an embedding character concatenation n , having its left neighbors $(r^{\text{char}} - 1)/2$, and its right neighbors $(r^{\text{char}} - 1)/2$: The word-level embedding for CNN is being represented by Eqs. 1 and 2 as follows.

$$x^n \left(k_{n-\frac{(r^{\text{char}}-1)}{2}}^{\text{char}}, \dots, k_{n+\frac{(r^{\text{char}}-1)}{2}}^{\text{char}} \right)^T \quad (1)$$

The vector $k^{\text{char}} \in R^{c_u^0}$ having j th element computed by convolutional layer that is the w embedding character level as follows:

$$[k^{\text{wchar}}]_j = \max_{1 < n < N} [w^0 x_n + a^0]_j \quad (2)$$

where

$$w^0 \in R^{c_u^0 \times d^{\text{char}}} \leftarrow \text{Convolution layer weight matrix}$$

The local feature is extracted using the same matrix around each window character in a given word.

4.4.4 Sentence-Level Scoring

A sentence x is given with m words $\{w^1, w^2, \dots, w^m\}$ and converting that to word-level joint and embedding-level character $\{u^1, u^2, \dots, u^n\}$, SCNNchar next step comprises of representation of sentence-level extraction r_x^{st} . A sentence-wide extraction of feature set method deals with two issues:

- Different sizes of sentences
- At any position in the sentence appears the important information.

These issues are tackled with the use of convolutional layer for computing the feature vector of wide sentence r^{st} . The convolutional neural network architecture, second convolutional layer, operates in a similar way as the layer is used for character-level feature extraction. Local features are produced in this layer around each word, and they are combined after which utilizes max operation for creating feature vector of fixed size for sentence.

A matrix-vector operation applied by second convolution layer for each window size of k^{word} in successive window sequence $\{u^1, u^2, \dots, u^n\}$. The vector $x_m \in R^{(d^{\text{word}}+c_u^0)k^{\text{word}}}$ as sequence concatenation of embedding k^{word} in n th word centralization: Sentence-level scoring is being represented by Eqs. 3 and 4.

$$x_m = (u_{m-\frac{(u^{\text{word}-1})}{2}}, \dots, u_{m+\frac{(u^{\text{word}-1})}{2}}) \quad (3)$$

The j th element is computed by convolutional layer having vector $r^{\text{st}} \in R^{c_u^1}$ as follows:

$$[r^{\text{st}}]_j = \max_{1 < m < M} [w_{x_m}^1 + a^1]_j \quad (4)$$

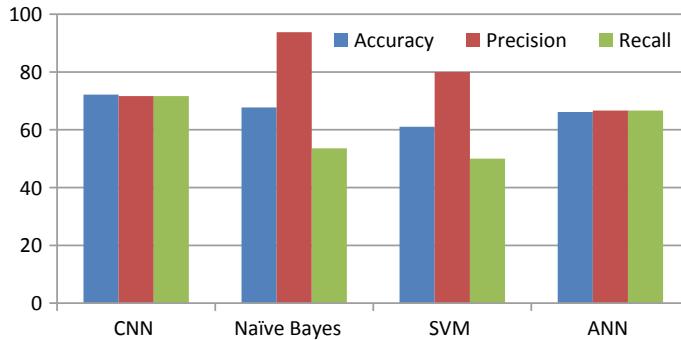
where

$$w^1 \in R^{c_u^1 \times (d^{\text{word}}+c_u^0)k^{\text{word}}} \leftarrow \text{Convolutional layer weight matrix.}$$

Finally, r_x^{st} vector having feature vector “global” of x sentence which is processed with the two neural network layer uses that extracts one more representation level, and each sentiment label $\tau \in T$ score is computed:

Table 1 Comparison table of CNN, Naïve Bayes, SVM, and ANN (accuracy, precision, and recall)

Classifiers	Accuracy	Precision	Recall
CNN	72.17	71.66	71.66
Naïve Bayes	67.72	93.75	53.57
SVM	61.06	80	50
ANN	66.17	66.66	66.66

**Fig. 3** Accuracy, precision, and recall comparison graph of different classifiers

$$S(x) = w^3 h(w^2 r_x^{\text{st}} + a^2) + a^3$$

where learning parameter matrix and vectors,

$$w^2 \in R^{h_u \times c_u^1}, \text{ and}$$

$$w^3 \in R^{|T| \times h_u}$$

$h(\cdot) \leftarrow$ Hyperbolic tangent.

5 Results and Discussion

The results of the performed experiment are being represented in Table 1:

The graphical representation of results of experiments is being represented in Fig. 3:

6 Conclusion and Future Work

This thesis presents an implementation of classification of sentiment analysis by aspects base ontology for semantic feature extraction and training the feature on four classifiers: CNN, ANN, naive Bayes, and SVM.

In experimental observation, important impact of ontology on information gathering of specific domain reduces the false-positive rate. In this thesis, comparison of precision, recall, and accuracy of different classifiers like SVM, ANN, Naïve Bayes, and CNN is shown. In experimental configuration, CNN shows significant difference of 72.17%, 71.66, and 71.66 in accuracy, precision, and recall, respectively. Based on the experimental setup, using ten cross-validation and convolution neural network shows high accuracy for a particular dataset for the positive and negative dataset.

An interested researcher may implement a few more things to this work for future study:

- In CNN, an important challenge is that how sources can be reduced? In this, enhancement is on the basis of effective data structure or parallel computing which speeds up the process.
- For future enhancement, our work on the semantic features optimization by meta-heuristic relative feature extraction can be presented.
- In the future, this work can be used in real applications and development system accordingly.

References

1. Kim, Y., Zhang, O.: Credibility adjusted term frequency: a supervised term weighting scheme for sentiment analysis and text classification (2014). arXiv preprint [arXiv:1405.3518](https://arxiv.org/abs/1405.3518)
2. Stojanovski, D., et al.: Twitter sentiment analysis using deep convolutional neural network. In: International Conference on Hybrid Artificial Intelligence Systems. Springer International Publishing, 2015
3. Ouyang, X., et al.: Sentiment analysis using convolutional neural network. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM). IEEE, 2015
4. Yang, T., et al.: Tb-CNN: joint tree-bank information for sentiment analysis using CNN. In: Control Conference (CCC), 2016 35th Chinese. IEEE, 2016
5. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data* **2**(1) (2015)
6. Gao, Y., et al.: Convolutional neural network based sentiment analysis using Adaboost combination. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016
7. Wu, H., et al.: Aspect-based Opinion Summarization with Convolutional Neural Networks. In: 2016 International Joint Conference on Neural Networks (IJCNN). IEEE, 2016
8. Poria, S., et al.: Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis. *Neurocomputing* (2017)
9. Jotheeswaran, J., Kumaraswamy, Y.S.: Opinion mining using decision tree based feature selection through manhattan hierarchical cluster measure. *J. Theor. Appl. Inf. Technol.* **58**(1) (2013)
10. Kaur, Jasakaran, Vashisht, Sheveta: Analysis and indentifying variation in human emotion through data mining. *Int. J. Comput. Technol. Appl.* **3**(6), 1963 (2012)
11. Ravichandran, M., Kulanthaivel, G.: Twitter sentiment mining (TSM) framework based learners emotional state classification and visualization for E-learning system. *J. Theor. Appl. Inf. Technol.* **69**(1) (2014)
12. Jindal, S., Singh, S., Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning. In: 2015 International Conference on Information Processing (ICIPI). IEEE, 2015

BrownBoost Classifier-Based Bloom Hash Data Storage for Healthcare Big Data Analytics



S. Arun Kumar and M. Venkatesulu

Abstract Healthcare big data analytics is the process of collecting and analyzing a large volume of patient data to find out useful information. Big data analytics has a number of challenges in many fields including cloud healthcare systems. The healthcare industry significantly creates large volumes of patient data. Most of the recent research works happen on big data analytics-enabled business models for increasing the prediction accuracy to reduce the risk level of patients. However, storage of data is a major concern, and data need to be accessed effectively across various locations in the distributed environment. Our aim is to develop a BrownBoost Classifier-Based Bloom Hash Data Storage (BBC-BHDS) mechanism for storing and accessing the healthcare data from various locations in distributed environment with minimum space usage and in less time. Initially, a large volume of data (i.e., patient data) are collected based on certain features (parameters), and then classified the input using BrownBoost Classification (BBC) algorithm. BrownBoost employs a non-convex potential loss function and uses the base SVM classifier for classifying the patient data. Experiment of the proposed BBC-BHDS mechanism is carried out on number of data files. The results show that the proposed BBC-BHDS mechanism is more efficient with respect to classification accuracy, false positive rate, space complexity, and data accessing time in comparison with existing methods.

Keywords Base SVM classifier · Bloom hash function · Data storage · Data analytics

S. Arun Kumar · M. Venkatesulu (✉)

Kalasalinagam Academy of Research and Education, Krishnankoil 626126, India
e-mail: m.venkatesulu@klu.ac.in

S. Arun Kumar
e-mail: arunkumar@klu.ac.in

1 Introduction

Big data analytics is used for analyzing the large volume of data to extract valuable information. Big data is significantly used in cloud healthcare applications to access patient information. For biomedical and healthcare communities, storing and accessing a large volume of data efficiently are a major concern. The big health data store and retrieve model employed in [1] depends on NoSQL databases for storage of healthcare data. However, the space complexity was unaddressed. Cloud-based MapReduce Model was developed in [2] for big data analysis, but the data classification was not performed. A high-performance computing (HPC) solutions in bioinformatics were developed in [3] to store and access the data, but the space complexity was not addressed. A distributed framework was developed in [4] for protecting patient data, but it failed to classify the data for extracting useful information. A big data analytics-enabled transformation model was developed in [5] based on practice-based view, which exposes the fundamental associations among big data analytics capabilities, but the transformation model did not perform efficiently on data storage. An efficient analysis of healthcare big data was carried out in [6] using a time distribution map investigation, but it has high time complexity. A cyber-physical system known as Health-CPS was introduced in [7] for patient-centric healthcare services, and it is also used for cloud and big data analytics. However, the classification of various patient data remained unsolved. Omic and EHR data characteristics with data analytics were performed in [8] with data preprocessing, mining, and modeling challenges. But the data accessing was not efficient and has large space complexity. Cloudlet-based mobile cloud computing framework was described in [9] for healthcare big data analysis, but it has high costs and failed to provide better clinical and operational processes. In [10], big data analytics was performed on healthcare data which was gathered from multiple sources to increase quality. A few issues are identified from the existing methods such as lack of classification, high space complexity, and problems with data storage and accessing. Hence, there is a need for schemes which handle both data storing and accessing capabilities efficiently. Following are the major contributions of the paper.

- BrownBoost Classifier-Based Bloom Hash Data Storage (BBC-BHDS) mechanism is described to store and access the big healthcare data. The patient's data is collected in a file format. Then, the BrownBoost Classifier is applied to perform effective classification of patient data with SVM classifier. The SVM classifier classifies the patient data into a number of groups. BrownBoost Classifier combines the weak hypothesis (i.e., weak classifier) to improve classification accuracy with less false positive rate.
- Bloom filter is used for storing the classified data in bit arrays using number of hash functions. A unique hash value for each patient data is generated and stored. As the data are stored using the hash value, it reduces the space complexity and access time.

The paper is organized as follows, Sect. 2 briefly discusses BrownBoost Classifier-based Bloom Hash Data Storage (BBC-BHDS) mechanism with a neat diagram. Section 3 discusses the experimental evaluation. Section 4 also presents analysis of the proposed BBC-BHDS mechanism with respect to various parameters. Section 4 presents comparison with the existing methods. Concluding remarks are summarized in Sect. 5.

2 BrownBoost Classifier-Based Bloom Hash Data Storage for Healthcare Big Data Analytics

Big data in healthcare industry is very huge and increasing exponentially. Cloud computing allows all big data operations through the provision of large storage. In healthcare industry, the patient data is complex in nature and very difficult to analyze. Therefore, the patient information (i.e., file) stored in health database keeps continuously increasing and considered as big data. The storage and accessing of patient data is a major challenge in healthcare environment. In order to improve storage and accessing performance of patient data, an efficient BrownBoost Classifier-Based Bloom Hash Data Storage (BBC-BHDS) mechanism is proposed in this paper. The architecture of BBC-BHDS mechanism is given in Fig. 1.

Figure 1 shows the BrownBoost Classifier-Based Bloom Hash Data Storage (BBC-BHDS) mechanism for storing and accessing the healthcare data with minimal space usage and less data accessing time. The proposed method involves three processing steps for accessing and storing the healthcare big data in efficiently. Initially, a large number of patient's data is collected. Then BrownBoost Classification (BBC)

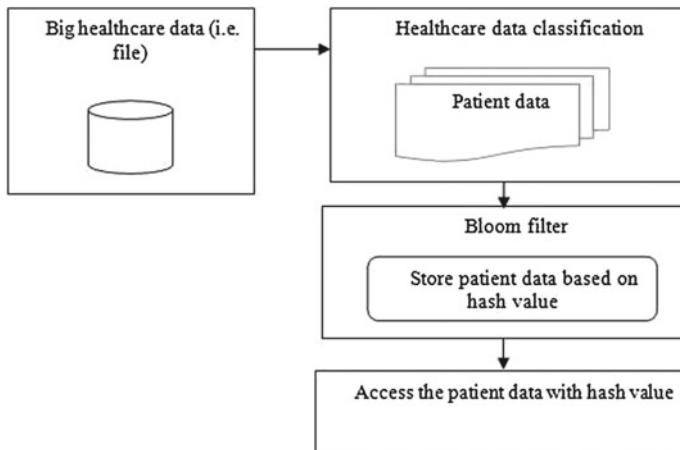


Fig. 1 Flow processing diagram of BrownBoost Classifier-Based Bloom Hash Data Storage mechanism

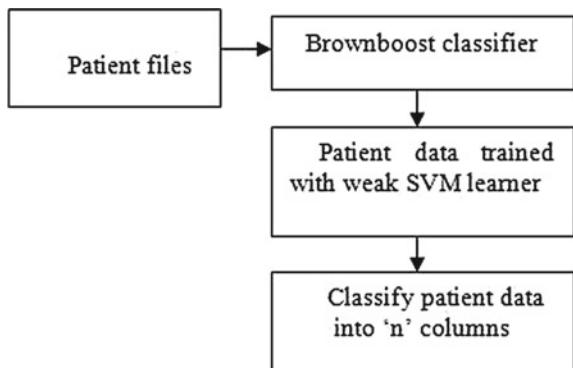
algorithm is used to classify with high accuracy the healthcare data based on physician requirement. Finally, Bloom hash function is used to store the classified data. Bloom filter is an efficient data structure used for storing and accessing the patient's data by means of hash value with minimal space usage and less data accessing time.

2.1 BrownBoost Classification Algorithm

Initially, a large number of patient data are collected and put into a file. Classification is performed using BrownBoost Classification (BBC) algorithm on the data. BrownBoost is a boosting and binary classifier. The conventional boosting technique does not perform well on noisy data sets, and the result of the final boosting classifier provides misclassification. Therefore, the BrownBoost Classifier effectively reduces the misclassifications. By using BrownBoost Classifier, the strong classifier combines weak hypotheses (i.e., weak learner). BBC-BHDS mechanism uses a weak learner as SVM classifier. Thus, BrownBoost makes a weak learner into a “strong” classifier through linear combination of weak hypothesis. A support vector machine (SVM) is a selective classifier which classifies the training samples (i.e., patient data) using a separating hyperplane. Therefore, BBC-BHDS mechanism combines the weak learner into strong using BrownBoost. The BrownBoost Classification process is described.

Figure 2 shows the BrownBoost Classification with weak SVM leaner performing efficient classification of patient data. SVM base classifiers are iteratively trained by using BrownBoost Classifier. Therefore, weak learner SVM is given to inputs of BrownBoost Classifier to obtain the strong classifier. BrownBoost Classifier contains ‘ n ’ number of training samples $(x_0, y_0), (x_1, y_1) \dots (x_n, y_n)$ where $x_i \in X$ is a set of input patient data and ‘ Y ’, the final classifier output, results as $Y_i = (-1, +1)$. SVM is a machine learning classifier which constructs optimal hyperplane in a high-dimensional data space for classifying the patient data. Considering the number of

Fig. 2 Process of BrownBoost Classification



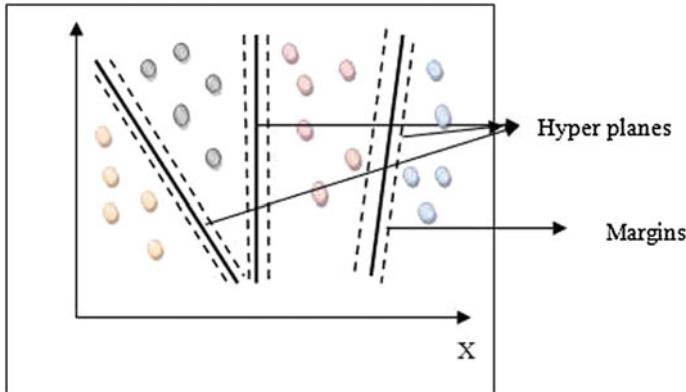


Fig. 3 Support vector machine

patient data $D = D_1, D_2, D_3 \dots D_n$ distributed in healthcare systems, classification of patient data using SVM classifier is shown in Fig. 3.

Figure 3 shows the support vector machine to classify the patient data accurately by using the optimal separating hyperplane which maximizes the margin of training data (i.e., patient data). The support vectors belonging to the margin that affects the direction of the discriminate hyperplane. At each occurrence, the patient data are classified on both the sides of the hyperplanes. From the figure, three hyperplanes are used to classify the patient data into ‘ n ’ number of columns (i.e., patient ID, age, sex, and so on). As a result, the different training data are generated for efficient classifications. An optimal hyperplane in SVM is described as the set of points which is denoted as

$$\vec{W} \cdot \vec{x} - \vec{b} = 0 \quad (1)$$

In step1, \vec{W} represents the weight vector and set of transiting data \vec{x} ($D = D_1, D_2, D_3 \dots D_n$) and \vec{b} denotes a bias. The region surrounded by the hyperplanes is called as margin. The margins of the hyperplane are expressed as,

$$\vec{W} \cdot \vec{x} - \vec{b} = +1 \quad (2)$$

$$\vec{W} \cdot \vec{x} - \vec{b} = -1 \quad (3)$$

The distance between the two margins is described as,

$$D = \frac{2}{\| \vec{W} \|} \quad (4)$$

In Eq. 4, ‘ D ’ represents a distance between two marginal hyperplane. The output of the weak Learner is described as follows,

$$Y = \sum_{i=1}^n W_i h_i(x_i) y_i \quad (5)$$

In (5), $Y \sum +1, -1$ denotes an output of SVM classifier. By using Eq. (5), SVM classifier classifies patient data into number of columns. BrownBoost Classifier is applied order to further improve the classification accuracy of the patient data. We recall that BrownBoost uses a non-convex potential loss function for efficient classification. The initial potential loss function is described as,

$$\Phi_L = \frac{1}{N} \sum_{i=1}^n 1 - e_r(\sqrt{t}) \quad (6)$$

In (6), Φ_L denotes a potential loss function and ‘ N ’ represents the number of patients data, ‘ t ’ denotes total time the algorithm is set to run, and e_r is the error functions. The final potential loss function is described as follows,

$$\Phi_L = \frac{1}{N} \sum_{i=1}^n 1 - e_r\left(\frac{R_i(x_i)}{\sqrt{c}}\right) \quad (7)$$

$$\Phi_L = 1 - e_r(\sqrt{c}) \quad (8)$$

In (7), R_i denotes margin of data, and ‘ C ’ denotes positive real value (parameter). The final classifier is a linear combination of weak hypotheses. Initially, the weight of each patient data is measured as follows,

$$W_i = \exp\left(-\frac{(R_i(x_i) + t_r)^2}{c}\right) \quad (9)$$

In (9), W_i represents weight of the patient data, R_i denotes margin of data, and t_r denotes remaining time. The weak learner is determined by normalizing the weight. The values of η_i and t_r are measured as follows,

$$\sum h_i(x_i) Y_i \exp\left(-\frac{(R_i(x_i) + \eta_i h(x_i) Y_i + t_r - t)^2}{c}\right) \quad (10)$$

In (10), η denotes an adjustment coefficient to ensure the final classification model. The margins for each data are updated as follows,

$$R_i(x_i + 1) = (R_i(x_i) + \eta_i h(x_i) Y_i) \quad (11)$$

In (11), $R_i(x_i + 1)$ denotes updated margins. The remaining time (t_r) is also updated as follows,

$$t_{r+1} = t_r - t \quad (12)$$

In (12), t_r represents remaining time, t denotes variable amount of time which is directly related to the weight given to the hypothesis. The output of final classifier is described as follows,

$$H(X) = \text{sign} \left(\sum_{i=1}^n \eta_i h_i(x_i) \right) \quad (13)$$

In (13), ' $H(X)$ ' denotes a strong classifier output to improve the classification accuracy. In Eq. (13), sign represents positive and negative results of output classifier. Therefore, the strong output classifier is described as follows,

$$H(X) = \begin{cases} +1 & \text{correctly classified} \\ -1 & \text{incorrectly classified} \end{cases} \quad (14)$$

In (14), $H(X) = +1$ denotes a correctly classified patient data, and $H(X) = -1$ denotes an incorrectly classified data.

In Fig. 4 describes Algorithm 1 describes a BrownBoost Classifier with SVM base learner for classifying the patient data effectively. Initially, the weight of patient data is measured. Next, the weak classifier is determined when the output of weak learner is greater than zero; the weak classifier is combined to provide efficient classification results. Finally, the value of adjacent coefficient and the amount of remaining time are derived, and the margin of each patient data and remaining time

Input: $\{(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)\}$ is a set of training samples, remaining time t_r , η adjustment coefficient, Output: Improve classification accuracy and reduce false positive rate Step 1: Begin Step 2: For each patient data Step 3: Initialize an amount of remaining time ' t_r ' Step 4: Calculate weight of each patient data using (9) Step 5: Find weak classifier such that $\sum w_i h_i(x_i) y_i > 0$ Step 6: Determine the value of η_i and t_r using (10) Step 7: Update the margins for each patient data using (11) Step 8: Update remaining time using (12) Step 9: if strong classifier output result $H(X) = +1$ then Step 10: patient data is correctly classified Step 11: else Step 12: patient data is incorrectly classified Step 13: End if Step 14: End for Step 15: End
--

Fig. 4 BrownBoost Classification algorithm

are updated. As a result, if the strong BrownBoost Classifier output +1 indicates the patient data are correctly classified and output –1 means the classifier is incorrectly classified. Hence, BrownBoost Classifier improves classification accuracy with less false positive rate.

2.2 Bloom Hash Function for Patient Data Storage

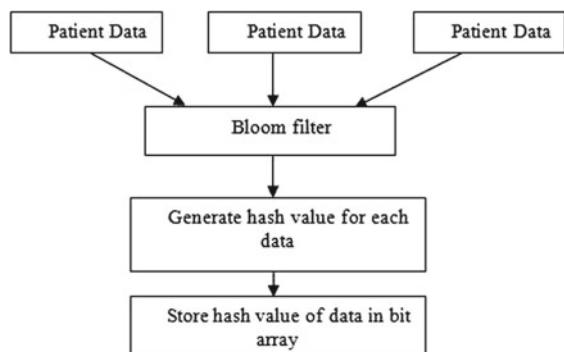
After classification, the data is stored using Bloom filter with the help of hash function. Bloom hash function is a probabilistic data structure used to store the patient data with minimum space complexity. Union and intersection of Bloom filters with a similar size and the number of hash functions are employed by using the logical operator, namely bitwise OR and AND, respectively. Union operation on Bloom filters is lossless, whereas intersection operation overcomes the false positive probability in the resulting Bloom filter. Bloom hash filter takes as input the patient data. A Bloom hash function takes input data of various sizes and outputs a fixed length string which is called as hash value.

Figure 5 shows the process of Bloom filter with the patient data as an input. The only way to reconstruct the input data from a Bloom hash function is to generate the correct hash value which is of fixed size. Use of large bit array in blooming filter concept efficiently reduces the false positive probability. Bloom filters stores classified patient data $D = D_1, D_2, D_3, \dots, D_n$ in an ' m_a ' bits of array using hash functions. The hash value of each patient data is denoted as $H = h_1, h_2, h_3, \dots, h_n$. Each data has different hash function which is stored in bit array.

Figure 6 shows the patient data is accessed from the Bloom filter with the hash value. The hash values are used to access the patient data from Bloom filter. This helps to effectively decrease the data access time. The Bloom hash function data storage algorithm is described as follows.

Figure 7 describes that Algorithm 2 describes the step-by-step process of Bloom hash function data storage with minimum accessing time. By using Bloom filter, the

Fig. 5 Process of Bloom filter



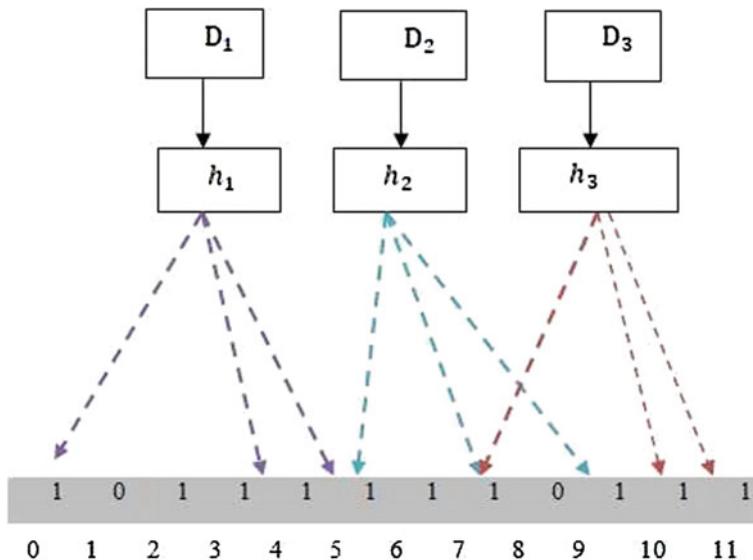


Fig. 6 Data access from Bloom filter

Input: Number of classified patient data $D_i = D_1, D_2, D_3, \dots, D_n$,

Output: Reduced space complexity with minimum access time

Step 1: Begin

Step 2: For each classified data D_i

Step 3: Generate the number of hash value using (21)

Step 4: Store the hash value to ' m_a ' bit array with fixed length using (24)

Step 5: If Physician request to access stored data with hash value then

Step 6: Requested Data is accessed

Step 7: end if

Step 8: end for

Step 9: End

Fig. 7 Bloom hash function data storage algorithm

classified data is stored in fixed size of array bits ' m_a ' using hash function ' r '. This helps to reduce the space complexity problem. Then, the physician can access the data along with hash value from the Bloom filter. Therefore, BBC-BHDS mechanism improves the data storage efficiency with minimum space complexity and access time.

3 Experimental Settings

Experimental evaluation of BBC-BHDS mechanism is carried out using Java Language. Experimental evaluation is performed to store and access the patient data from the file. For the experimental consideration, healthcare patient data are taken from openly available UCI machine learning repository. In our simulation, heart diseases patient's data sets are considered for storing and accessing the patient data from the file in an effective manner in minimum time. The database contains 76 attributes for healthcare big data analytics. The number of attributes used in patient's record files is patient identification number, age, sex, and so on. There are 303 instances used in this data set. The data were collected from Cleveland Clinic Foundation. Experimental evaluation using BBC-BHDS mechanism is conducted on different factors such as classification accuracy, false positive rate, space complexity, and data accessing time.

4 Results and Discussions

Experimental results of BBC-BHDS mechanism is conducted with respect to different parameters such as classification accuracy, false positive rate, space complexity, and data accessing time and compared with the existing methods, namely big health data store and retrieve model [1] and Map Reduce Model [2]. The performance of BBC-BHDS mechanism is evaluated using number of files. The results of BBC-BHDS mechanism is discussed with help of tables and graph values.

4.1 Impact of Classification Accuracy

Classification accuracy is defined as the ratio of the number of (i.e., no. of) patient data correctly classified to the total number of files. The mathematical formula for classification accuracy is described as follows,

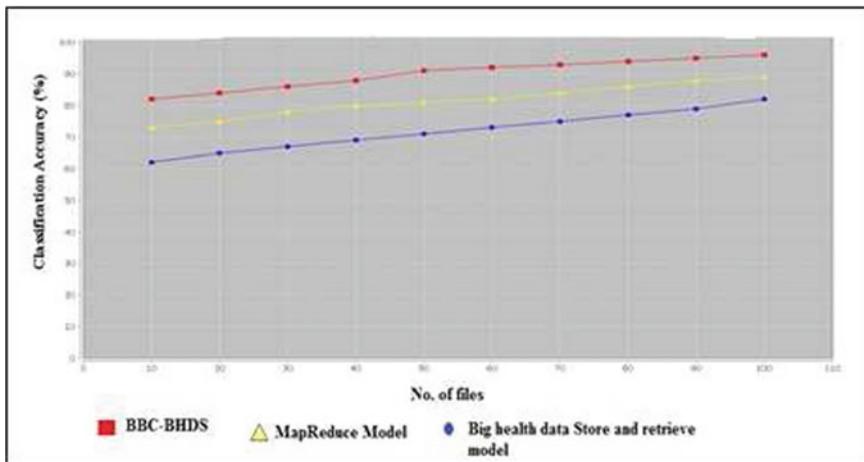
$$CA = \frac{\text{No. of patient data correctly classified}}{\text{No. of files}} * 100 \quad (15)$$

In (15), CA denotes classification accuracy and it is measured in terms of percentage (%). Each file has individual patient data for efficient data processing.

Table 1 shows experimental results of classification accuracy with respect to number of files which contains the patient data. The accuracy is measured based on the correctly classified number of columns using the classification performed on the patient data. The classification accuracy has considerably increased using BBC-BHDS mechanism when compared to the existing big health data store and retrieve model [1] and Map Reduce Model [2]. Figure 8 shows the performance of classification accuracy with respect to number of files. Each file has different patient data

Table 1 Tabulation for classification accuracy

No of files	Classification accuracy y (%)		
	BBC-BHDS	Big health data store and retrieve model	Map Reduce Model
10	82	62	73
20	84	65	75
30	86	67	78
40	88	69	80
50	91	71	81
60	92	73	82
70	93	75	84
80	94	77	86
90	95	79	88
100	96	82	89

**Fig. 8** Measure of classification accuracy

which is used for classification. As shown in Fig. 8, the classification accuracy is significantly more in using BBC-BHDS mechanism compared to the existing methods. Initially, the input files contain patient data. The data are classified into number of columns by applying the SVM base classifier. The SVM classifier classifies the patient data using separate hyperplane which maximizes the margin of training data (i.e., patient data). The data belonging to the margin of the discriminate hyperplane are classified into ' n ' number of columns such as patient ID, age, sex, and so on. The BrownBoost Classifier is applied to combine weak learner to make a strong classification results. By using BrownBoost Classification approach, the weak classifier is determined when the output of the classifier is greater than zero. Then, the margin

of each patient data and remaining time are updated. This process is repeated until the entire patient data are correctly classified into number of columns. As a result, classification accuracy has increased by 10 and 25(%) when compared to the existing big health data store and retrieve model [1] and Map Reduce Model [2], respectively.

4.2 Impact of False Positive Rate

False positive rate is defined as the ratio of number of patient data incorrectly classified to the total number of files. Mathematical formula for false positive rate is described as follows,

$$FPR = \frac{\text{No. of patient data incorrectly classified}}{\text{No. of files}} * 100 \quad (16)$$

In (16), FPR denotes a false positive rate and is measured in terms of percentage (%). Lower the false positive rate more efficient the method is said to be.

In Table 2, analysis of false positive rate with respect to number of files is considered. Table shows that the false positive rate of three different methods BBC-BHDS mechanism and existing big health data store and retrieve model [1] and Map Reduce Model [2]. The false positive rate of the proposed BBC-BHDS mechanism is significantly reduced when compared to the other two methods. Figure 9 depicts the performance analysis of false positive rate with respect to number of input files. The figure clearly shows that the proposed BBC-BHDS mechanism effectively classifies the number of patient data with less false positive rate compared to the existing methods. During each iteration, the BrownBoost Classifier reduces the misclassifi-

Table 2 Tabulation for false positive rate

No of files	False positive rate (%)		
	BBC-BHDS	Big health data store and retrieve model	Map Reduce Model
10	20	40	31
20	22	42	33
30	24	44	35
40	26	46	36
50	28	48	38
60	30	50	39
70	31	51	41
80	33	52	43
90	34	53	45
100	35	54	47

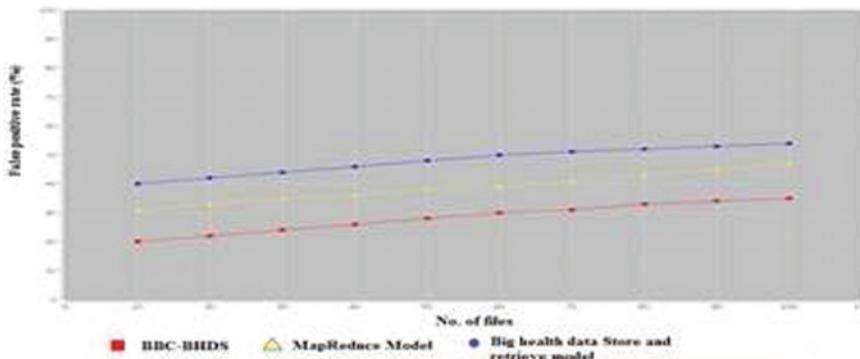


Fig. 9 Measure of false positive rate accuracy

cation by combining the base classifier. If the output of strong classifier provides the negative results, then the patient data are incorrectly classified. But the proposed BBC-BHDS mechanism improves classification accuracy with less false positive rate by avoiding the misclassification. Considering the input of 10 files, the false positive rate of the proposed method is 20(%), whereas big health data store and retrieve model and Map Reduce Model are 40 and 31(%), respectively, which show significant improvement of BBC-BHDS mechanism. The false positive rate is reduced by 42 and 28(%) using MLABC technique when compared to big health data store and retrieve model [1] and Map Reduce Model [2], respectively.

4.3 Impact of Space Complexity

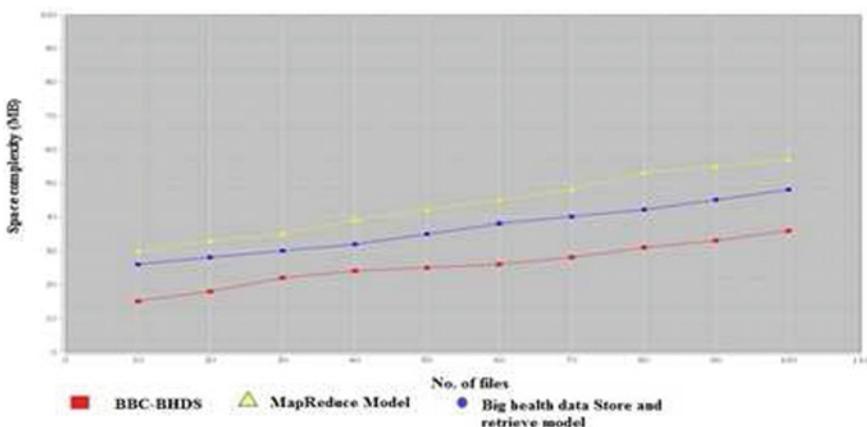
Space complexity refers to amount of storage space required to store the classified patient data from the files using Bloom filter. The space complexity is measured in terms of megabytes (MB). The space complexity is measured as follows,

$$SC = \text{No. of files} * \text{Space (Storing patient data)} \quad (17)$$

Lower the space complexity more efficient the method is said to be. Table 3 shows the performance result of space complexity with respect to number of input files. The space complexity is measured based on the amount of storage space consumed for an algorithm to store the classified patient data. From the table value, the space complexity using BBC-BHDS mechanism is considerably reduced when compared to big health data store and retrieve model [1] and Map Reduce Model [2]. Figure 10 illustrates space complexity versus number of files. The classified patient data are stored effectively for further processing. The space complexity during the data storage is reduced using BBC-BHDS mechanism than the existing methods. It is so because BBC-BHDS mechanism uses Bloom filter for efficient data storage. The

Table 3 Tabulation for space complexity

No of files	Space complexity (MB)		
	BBC-BHDS	Big health data store and retrieve model	Map Reduce Model
10	15	26	30
20	18	28	33
30	22	30	35
40	24	32	39
50	25	35	42
60	26	38	45
70	28	40	48
80	31	42	53
90	33	45	55
100	36	48	57

**Fig. 10** Measure of space complexity

Bloom filter generates the hash function for each classified data and is stored in a bit array. Therefore, the Bloom hash function is used to reconstruct the patient data in a file to generate a fixed size of hash value. Therefore, the BBC-BHDS mechanism improves the data storage with minimum space complexity. As a result, the space complexity is reduced by 30 and 41(%) when compared to big health data store and retrieve model [1] and Map Reduce Model [2], respectively.

Table 4 Tabulation for data accessing time

No of files	Data accessing time (ms)		
	BBC-BHDS	Big health data store and retrieve model	Map Reduce Model
10	15	22	28
20	18	25	30
30	20	28	35
40	22	31	38
50	24	33	42
60	25	34	45
70	27	37	48
80	28	40	50
90	30	42	52
100	32	45	54

4.4 Impact of Data Accessing Time

Data accessing time is defined as the amount of time required to access patient data from a file. The formula for data accessing time is described as follows,

$$\text{DAT} = \text{No. of files} * \text{Time (Access the data)} \quad (18)$$

In (18), ‘DAT’ denotes the data accessing time and it is measured in terms of milliseconds (ms).

Table 4 describes data accessing time with different methods namely BBC-BHDS mechanism, big health data store and retrieves model [1], and Map Reduce Model [2]. The data accessing time is considerably reduced by using BBC-BHDS mechanism when compared to the existing methods. Figure 11 depicts data accessing time with respect to number of files. As shown in Fig. 10, data accessing time is considerably reduced using BBC-BHDS mechanism than the existing methods.. In BBC-BHDS mechanism, patient data are stored in the form of n number of columns. Then, physician requests to access patient data with the corresponding hash value and improves accessing of patient data with minimum time. Considering 10 input files, data accessing time of BBC-BHDS mechanism is 15 ms whereas 22 and 28 ms while using big health data store and retrieve model [1] Map Reduce Model [2], respectively. Data accessing time is reduced by 29 and 43(%) when compared to big health data store and retrieve model [1] Map Reduce Model [2], respectively. The above results show that the proposed BBC-BHDS mechanism can be effectively used for storing and accessing the healthcare data with minimal space complexity and less data accessing time.

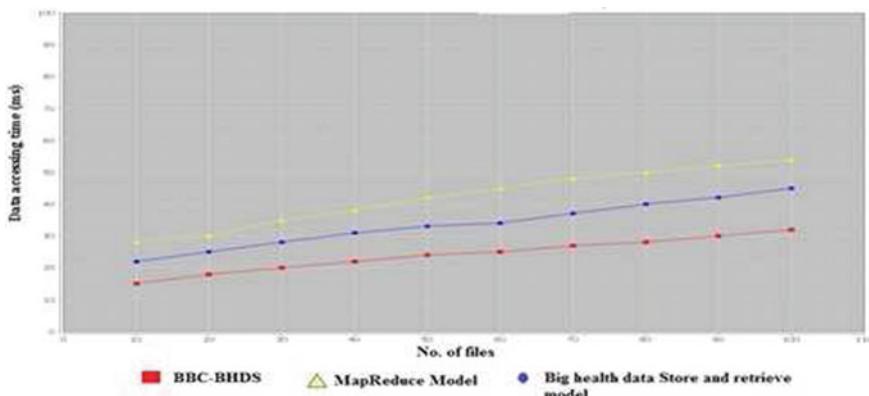


Fig. 11 Measure of data access time

5 Conclusion

An efficient mechanism called BrownBoost Classifier-based Bloom Hash Data Storage (BBC-BHDS) is introduced for big healthcare analytics in cloud. Initially, the number of patient data is collected and arranged in a file. A file consists of number of patient information. Then, the BrownBoost Classification (BBC) is performed to classify the patient data into number of columns such as patient ID, age, and so on. This helps to improve the classification accuracy with less false positive rate. The classified data are stored in a bit array using Bloom hash function. The physician can access the data with corresponding hash value. Therefore, BBC-BHDS mechanism performs efficient storing and accessing the patient data in an effective manner. Experimental evaluation is conducted using heart diseases data set taken from UCI machine learning repository with the parameters such as classification accuracy, false positive rate, space complexity, and data accessing time. BBC-BHDS mechanism improves classification accuracy with minimum false positive rate, space complexity, and data accessing time than the state-of-the-art methods.

Acknowledgements The first author is thankful to the management of Kalasalingam Academy of Research and Education for providing fellowship.

References

- Goli-Malekabady, Z., Akbari, M.K., Sargoaei-Javan, M.: An effective model for store and retrieve big health data in cloud computing. *Comput. Methods Program. Biomed.* **132**, 75–82 (2016). Elsevier
- Sahoo, P.K., Mohapatra, S.K., Wu, S.L.: Analyzing healthcare big data with prediction for future health condition. *IEEE Access* **4**, 9786–9799 (2016)

3. Merelli, I., Perez-Sanchez, H., Gesing, S., D'Agostino, D.: Managing, analysing, and integrating big data in medical bioinformatics: open problems and future perspectives. *BioMed Res. Int.* **2014**, 1–13 (2014). Hindawi Publishing Corporation
4. Sarkar, B.K.: Big data for secure healthcare system: a conceptual design. *Complex Intell. Syst.* **3**(2), 133–151 (2017). Springer
5. Wang, Y., Kung, L.A., Yu, W., Wang, C., Cegielski, C.G.: An integrated big data analytics-enabled transformation model: application to health care. *Inf. Manag.* **55**, 1–16 (2017). Elsevier
6. Gu, D., Li, J., Li, X., Liang, C.: Visualizing the knowledge structure and evolution of big data research in healthcare informatics. *Int. J. Med. Inf.* **98**, 22–32 (2017). Elsevier
7. Zhang, Y., Qiu, M., Tsai, C.W., Hassan, M.M., Alamri, A.: Health-CPS: healthcare cyber-physical system assisted by cloud and big data. *IEEE Syst. J.* **11**(1), 88–95 (2017)
8. Wu, P.Y., Cheng, C.W., Kaddi, C.D., Venugopalan, J., Hoffman, R., Wang, M.D.: Omic and electronic health record big data analytics for precision medicine. *IEEE Trans. Biomed. Eng.* **64**(2), 263–273 (2017)
9. Tawalbeh, L.A., Mehmood, R., Benkhelifa, E., Song, H.: Mobile cloud computing model and big data analysis for healthcare applications. *IEEE Access* **4**, 6171–6180 (2016)
10. Batarseh, F.A., Latif, E.A.: Assessing the quality of service using big data analytics: with application to healthcare. *Big Data Res.* **4**, 13–24 (2016). Elsevier
11. Wang, Y., Kung, L.A., Byrd, T.A.: Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. *Technol. Forecast. Soc. Change* **126**, 1–11 (2016). Elsevier
12. Wang, Y., Hajli, N.: Exploring the path to big data analytics success in healthcare. *J. Bus. Res.* **70**, 287–299 (2017). Elsevier
13. Ma, Y., Wang, Y., Yang, J., Miao, Y., Li, W.: Big health application system based on health internet of things and big data. *IEEE Access* **5**, 7885–7897 (2017)
14. Wu, J., Lib, H., Liua, L., Zhenga, H.: Adoption of big data and analytics in mobile healthcare market: an economic perspective. *Electron. Commer. Res. Appl.* **22**, 24–41 (2017). Elsevier
15. Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. *Health Inf. Sci. Syst.* **2**(3), 1–10 (2014). Springer
16. Rallapalli, S., Gondkar, R.R., Pavan, U., Ketavarapu, K.: Impact of processing and analyzing healthcare big data on cloud computing environment by implementing hadoop cluster. *Proc. Comput. Sci.* **85**, 16–22 (2016). Elsevier
17. Ullah, F., Habib, M.A., Farhan, M., Khalid, S., Durrani, M.Y., Jabbar, S.: Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. *Sustain. Cities Soc.* **34**, 90–96 (2017). Elsevier
18. Mao, H., Sun, Y.: A way to understand inpatients based on the electronic medical records in the big data environment. *Int. J. Telemed. Appl.* **2017**, 1–9 (2017). Hindawi
19. Youssef, A.E.: A framework for secure healthcare systems based on big data analytics in mobile cloud computing environments. *Int. J. Ambient Syst. Appl.* **2**(2), 1–11 (2014)
20. Sun, J., Jeliazkova, N., Chupakhin, V., Golib-Dzib, J.F., Engkvist, O., Carlsson, L., Wegner, J., Ceulemans, H., Georgiev, I., Jeliazkov, V., Kochev, N., Ashby, T.J., Chen, H.: ExCAPE-DB: an integrated large scale dataset facilitating big data analysis in chemogenomics. *J. Cheminform.* **9**(17), 1–9 (2017). Springer

AmbuPod a Family-Owned Social Enterprise for Inclusive Quality Health care in India



Pradnya Vishwas Chitraq, Pravin Kumar Bhoyar and (Brig) Rajiv Divekar

Abstract India is one of the emerging economies in the twenty-first century. It is witnessing rapid progress in different sectors. At the same time, this progress is uneven and is not inclusive. There are many in India who do not get adequate food to eat. Many children do not get proper education. Many girl children and women are seen as liabilities and are side tracked in terms of nutrition, educational and career opportunities. Health care in India is one of the main areas where we see on the one hand a lot of sophisticated healthcare facilities available to a privileged few whilst the poor and the rural areas either have no access or find them too expensive. Some of them also get caught in lifelong debts on account of trying to procure medical attention for themselves or their family members. The twenty-first century is a witness to a lot of social enterprises that try to solve the problem of access to affordable quality health care for the poor and the needy. Start-ups and social enterprises are critical for India to grow in an inclusive and sustainable manner in today's economy. Social enterprises are said to meet two conditions—they address long-standing social problems and they develop innovative solutions to do so (Santos in *J. Bus. Ethics.* 111:335–351, 2012) [1]. They attempt to restore the balance in the social, structural and political systems by producing and sustaining positive social change. Such organizations try to make available goods and services that the market or public sector is either unwilling or unable to provide. They are engaged in developing skills, creating employment and integrating socially excluded people (Trivedi and Stokols in *J. Entrep.* 20(1):1–32, 2011) [2]. Quite a few entrepreneurial ventures in India are family owned wherein family members are partners. Some of the social enterprises are also family-owned especially at the start. The paper studies how a family-owned social enterprise namely AmbuPod has with the help of technology sought to innovatively make available affordable health care to the poor and rural people. The researchers have used the primary research technique of personal interview of a member of the family business. They have also referred to secondary data. The study is very important as it will encourage policy makers, NGOs and corporate to support social enterprises like

P. V. Chitraq (✉) · P. K. Bhoyar · (Brig) R. Divekar
Symbiosis Institute of Management Studies (SIMS), Constituent of Symbiosis International
(Deemed) University, Pune, India
e-mail: pradnyac@sims.edu

AmbuPod. The study also reveals how such social enterprises along with making available critical services like affordable health care, also provide employment.

Keywords Social enterprises • Family business • Technology • Affordable health care • Employment

1 Introduction

India is the world's second most populous country in the world, and the fifth largest economy in terms of GDP. More than three-fifths of India's population lives in rural areas [3]. Approximately, 53% of the country's total employment is in agriculture [4], yet only 19% of India's GDP is created by this sector. India ranks 134 among 187 countries in the world on the Human Development Index (HDI) as per HDI Report 2011 [5]. India has a lot of developmental issues like poverty, health care, education, unemployment, water and sanitation, environmental issues and livelihood. The growth of the Indian economy excludes vast masses of people who are also forced to bear the cost of this growth [6].

Social entrepreneurship identifies a social problem and then uses entrepreneurial principles to start and run a social venture in order to arrive at a desired social change. A lot of social entrepreneurs have emerged in India who have used innovative approaches to provide sustainable solutions to social problems, thereby creating value for the beneficiaries. Whilst a business entrepreneur typically measures performance in profit and return, a social entrepreneur also measures positive returns to society. In India, a social entrepreneur is usually the founder, co-founder or a chief functionary of a social enterprise, or a non-profit, which raises funds through fund raising events and community activities and occasionally products. Social entrepreneurship is becoming very prominent in India in the twenty-first century as the country is trying to achieve a balance between a growing GDP growth, ensuring inclusive growth and attempting to address issues ranging from education, inclusive health care, and energy efficiency to climate change.

The role of social enterprises in healthcare-related services is especially crucial. Many rural areas do not have hospitals or sometimes even ambulances. Poor people in cities also cannot afford these services. Critical time is wasted trying to get the person needing medical attention to the hospital. Some rural areas do not even have proper roads. People often ignore what are perceived to be ordinary cough and cold, and rely on home remedies. In some undeveloped villages, expecting mothers on the verge of delivering often have to make do with a midwife who may sometimes be engaged elsewhere or may not reach the patient on time. In fact, reaching social entrepreneurship to India's hinterland is a serious challenge [7].

The researchers studied the various social enterprises that try to make quality health care available to poor and rural areas through secondary data. They also studied one family initiated and owned enterprise called LYNK AmbuPod Pvt. Ltd. The enterprise is in its nascent stage and tries to in a very economical and technologically

efficient manner seeks to administer medical aid to the patient whilst transporting to the nearest possible hospital. It also seeks to dispense medical treatment on an OPD basis to rural and poor people on a day-to-day basis. The researchers have tried to find out how this social enterprise seeks, with the help of technology, to bring about quality health care in India in a cost-effective, ingenious and inclusive manner.

2 Literature Review

India is facing a lot of developmental issues like poverty and poor health care and unemployment. The reasons for these are the higher growth rate of population, constantly growing social and environmental needs and limited resources. Agriculture is still the major source of livelihood for many in the Indian subcontinent even though the economy has transformed because of the development of industries and service sectors during the years of planning in the post-independence era [8]. Poor people suffer a disproportionate share of the world's health problems. Extreme poverty causes ill health by forcing people to live in environments that make them sick, without access to decent shelter, clean water or adequate sanitation [9].

The concept of 'social entrepreneurship' has emerged as a global phenomenon in order to bridge the increasing gap of the demand for social and environmental needs and the corresponding supply of resources. Entrepreneurship aimed at economic development has so far received a lot of scholarly interest. But it is only recently that social entrepreneurship as a means for promoting social progress has attracted the attention of researchers [10].

The term 'social entrepreneurship' was coined by Ashoka Innovators for the Public in the 1980s. It is viewed as a solution to market and state failure in fulfilling the needs of society [11]. Social entrepreneurs today are recognized today as change agents in the social sector; they try to bring in systematic solutions to social problems, and thereby try to implement social change that creates social value [12]. With population growth, the demand for social needs is also continuously increasing, because people across the world are encountering similar problems such as inadequate education and healthcare systems, environmental threats, poverty and high crime rates.

India's vibrant SE space is young in terms of years of operation, and nascent in terms of revenue size per enterprise [13]. According to the *Beyond Profit* 2010 survey, about 68% of SEs have been in existence for five years or less. SE revenues are growing rapidly; for instance, nearly one-third of the enterprises surveyed by *Beyond Profit* grew by over 50% between 2009 and 2010, whilst only 6% of the surveyed enterprises had negative growths. SEs that adopt innovative business models with for-profit entities account for three-fifths of all SEs.

The definition of social entrepreneurship has changed over time. From corporate philanthropy to non-profit and now to self-sustainability, social entrepreneurship has evolved with time and the needs of the world. Today, mainstream financial institutions are actively involved in social entrepreneurship. Various venture capital firms are investing in for-profit entities with social objectives. Specialized social investors

provide capital, networking, marketing and business expertise to such ventures. As per Beyond Profit Survey, forty-five percent of respondents obtained funds from commercial sources, whereas 21% of respondents source their funds from personal connections such as family members and friends; another 21% rely on grants and donations from charitable organizations. Intellecap in its study [14: 9] defines a social enterprise as fulfilling four criteria namely:

- they are to be for-profit,
- are committed to social impact,
- they have a base of the pyramid (BOP) focus, and
- they serve a critical-needs sector.

This framing is not necessarily shared by many social entrepreneurs who have often engaged with the market and yet are reluctant to articulate their work in terms of social businesses.

2.1 Challenges that Social Enterprises Face

A social entrepreneur should have a clear understanding of the problems and issues they may face so that they can make informed decisions. Lack of capital is a major challenge for the Indian entrepreneur. Generally, the social entrepreneurs run their business with their own funds social enterprises have to face a hostile reaction from financial institutions and governments as far as funding is concerned. This forces social entrepreneurs to take, what can be, a more difficult path of approaching venture capitalist and philanthropic organizations.

Social entrepreneurs do not always work in a lucrative market; they identify a problem within society and try to find affordable solutions for them. Once they find the way to earn some profit after providing the best low-cost solution to the needs of the society, more traditional businesses tend to enter the market with a similar solution and technique. This increases the transaction costs and competition for social entrepreneurs, and thereby hampers their future growth.

Lack of government support is a major hindrance for social business development in India. As of now, the government is not promoting in any way these social cause ventures. Very complex and strict government's policies and regulations for social entrepreneurs and the absence of tax incentives or subsidies for a social business, serve as major impediment to the growth of social businesses in India [15].

Social enterprises, in order to fulfil their mission in a holistic manner, have to typically employ manpower from the underprivileged sector of the society, leading to increased training and developmental cost as these people are usually uneducated and unskilled. The challenge for social entrepreneurs is to enlist the support of foundations and philanthropists. Foundations and high-net-worth individuals are in a position to bring about social change, as they are free of forces that dominate the decisions of governments and business, respectively.

2.2 Social Enterprise Models

The early social enterprise models were basically framed on philanthropic models where charities and not-for-profits were constantly trying to generate income from donations and grants. But the new generation of social enterprise organizations are becoming self-sustaining for the long term rather than having to go cap in hand to donors every few years. Most of these foundations are keen on offering fellowships or grants to those non-profits that have been around for a long period or have been running their operations for a few years. Thus, the investment climate for social enterprises in India is yet to reach maturity. Investors, who decide to invest in social ventures, understand the blended value of a social enterprise and are not motivated merely by the financial returns. Impact investors believe that such investments are more sustainable than a commercial investment that would mainly focus on financial returns. In India, such an optimistic investment climate for social investments is yet to materialize.

Social enterprises help one get into a social impact mindset. One constantly starts thinking about constant profit and efficiency and optimization. This in turn leads to a lot of innovation, which benefits both the company and society. Korsgaard and Anderson [16] argued that entrepreneurship is enacted in a socialized context and produces social outcomes. They used ‘creation of social value’ as a means for understanding outcomes. Singh [17] has emphasized the social value creation dimension in her model of social enterprise which is depicted below (Fig. 1).

The model basically talks of five major dimensions:

- the capabilities of social entrepreneur/s who initiate the process of social value creation;
- the social value which they create;
- identification of opportunity;
- resource mobilization, i.e. ability to work without regard to resources currently controlled; and
- the context/environment, within which they operate.

Since the most important dimension is social value creation, it occupies a central position in the model. The other dimensions are directed towards the central position. The overlapping (interplay) between these dimensions is also reflected in the process of social value creation to achieve social mission.

Polonsky and Grau [18] defined ‘social value’ as the total social impact a charitable organization has on all its stakeholders. In their words, ‘The social value of charities can be defined as the charity’s effect on all of its stakeholders. These stakeholders include donors, employees, volunteers, other charities and non-profits, those the charity is helping, and society in general’ (p. 1 33). Where entrepreneurship is concerned with profitable opportunities, in social entrepreneurship, opportunities are identified and exploited for solving social problems, meeting social needs, creating social value and bringing about social change [19]. Value capture from an activity happens when the focal actor is able to appropriate a portion of the value created

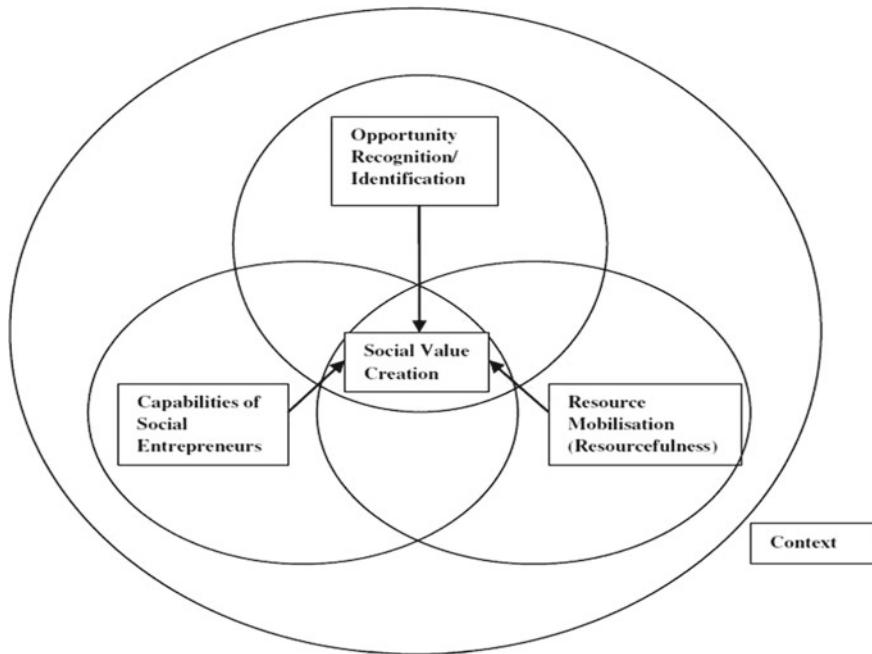


Fig. 1 Social entrepreneurship model: from Singh [17]

by the activity after accounting for the cost of resources that he/she mobilized [20]. Some level of value capture is important to ensure the growth and sustainability of the organizations whose activities create value. In fact, SEs that adopt innovative business models with for-profit entities account for three-fifths of all SEs [21]. For-profit models also include collective ownership structures such as cooperatives and producer companies. Models for social entrepreneurship in India are social for-profit enterprise, non-profit and hybrid model. Other ways of creating impact in India are through philanthropy and through corporate social responsibility.

3 Objective

The study seeks to find out how a social enterprise with the help of technology can, in an innovative manner, make available affordable health care to poor and rural people. It also seeks to study how such an enterprise can provide some training and meaningful employment to some people.

4 Research Methodology

The researchers interviewed the co-owner and co-founder of LYNK AmbuPod Pvt. Ltd. namely, Ms. Yamini Lavanian. They also referred to secondary sources of information

5 Findings

5.1 *Genesis of LYNK AmbuPod*

The researchers found in their structured interview of Ms. Yamini Lavanian that the idea behind the social enterprise LYNK AmbuPod Pvt. Ltd. was her father, Dr. Lavanian. He is a medical doctor from AFMC, Pune. He is a retired Air Force Officer. After retirement, he got into healthcare informatics, hospital information systems. In short, he started providing consultancy and training in healthcare IT. For some time, he was VP of telemedicine at Apollo Hospital, Andhra Pradesh. There he realized that there is a big gap in rural areas because of expensive infrastructure that is unaffordable there. He started researching how technology can be roped into making healthcare infrastructure affordable and accessible to all. This research was financed out of his pension money. He finally came up with AmbuPod, a mobile ambulance cum OPD unit.

5.2 *Designing AmbuPod*

The engineering team led by Dr. Lavanian Dorairaj and based on his Patent (Pending) design, devised a narrow vehicle that would go through very small and hilly paths. The team designed a vehicle which was three feet in width (which is more than an air passenger seat) and which enables one medical assistant to give resuscitative help to a single patient who can be laid down flat. This is a typical anesthesiologist position wherein the patient's head is in the lap of the doctor. The doctor is in a position to reach out to all the vital organs till the torso. If a patient has a lower limb injury, AmbuPod will reach the place and first stabilize the patient, and then transfer to AmbuPod in order to take to a nearby hospital.

The kind of medical equipment kept in AmbuPod took the longest time for research. The following things had to be kept in mind regarding the equipment in the AmbuPod:

- The weight and portability of the equipment
- The quality had to be CE certified

- The size and amount of energy the equipment takes to run on solar energy (the batteries are charged by the solar panels. The batteries can also be charged from household electric connections.

AmbuPod has a suction machine and an oxygen tank. The team had to ensure that the tank would have adequate oxygen and yet would not be too heavy. A suction machine is needed if a patient is bleeding or choking on own saliva. It can be put in patient's mouth. The positioning of every equipment is crucial to save the patient in time. A lot of the equipment had to be bought from abroad. But some machines have been innovatively designed in India itself. For example, the 12 lead ECG machine is the size of a visiting/credit card. It gives accurate readings onto a tablet or an android phone. So if a normal ECG costs around Rs. 150 in a government hospital, AmbuPod offers it from Rs. 30 to Rs. 50 to rural people.

AmbuPod also offers mother and child care. It has a foetal Doppler which monitors the heart rate of the foetus. Many miscarriages and deaths happen in villages because they cannot find out the health of the foetus in the womb. AmbuPod takes care of this and prevents miscarriages with timely intervention based on this test. AmbuPod therefore is not just a low-cost single patient ambulance. It is also a mobile clinic that runs daily OPD services. It is a healthcare platform that is also telemedicine enabled.

5.3 Business Plans/Models Offered by the Enterprise

AmbuPod has a few business plans. One of them is that an AmbuPod be bought by village mukhiyahs or panchayat heads or NGOs. One AmbuPod can take care of three to five villages. AmbuPod goes to a village and selects one person from that village itself and recruits and trains that person. This person becomes the eyes and ears of the doctor. Dr. Lavanian has created a telemedical app called AmbuApp. This app takes down all the medical details of the patient which are logged in by the trained person who is designated as the rural manager (RM). A portable tent is opened where the AmbuPod stands. Folding chairs are set up for villagers. The moment the RM clicks the submit button of the app, all the details come to doctors in Pune; Sometimes, it is even possible to have video conferences as the app is video-enabled and can run on 2 GB (as villages have poor network). In case, video is not possible, consultation can be given by Pune doctors over phone to the RM who then dispenses the treatment given by way of an e-prescription issued by the doctor.

This kind of communication happens only if there is no Internet. If Internet connection is available, there is a whole seamless communication between the doctor and the RM and the patients via the AmbuApp. AmbuPod pays the RM, the doctors, nurses or the para medics. The RM and other staff are on a monthly salary and also get a commission if certain targets are exceeded per month. Usually, the paramedic who runs the AmbuPod is also the RM, and is the heart and soul of the mobile clinic. The patients are treated on a pay per use model wherein each consultation is for Rs. 75 only.

The owners have in mind other models as well. One of them is the PPP model namely, the public–private partnership, where the owners will collaborate with government to give affordable health care. Another model is the JV model wherein they will collaborate with strategic partners like corporate for their CSR. The owners also have the NGO model of business wherein they will periodically tie up with Rotary Club and other NGOs who purchase the AmbuPod and all services of that AmbuPod are run by the owners.

Still another model of business that the owners have is the donor model. Non-resident Indians (NRIs) and rich people can buy an AmbuPod and donate it to a village. Again, the owners will run operate it. In short, in all the business models offered, the owners will run the AmbuPod. Currently, their business model will generate employment among rural and needy people. They are also offering employment to trained, retired Armed Medical corp (AMC) personnel employment because of their skill and experience.

5.4 Strengths of the Enterprise

- AmbuPod is capable of conducting blood, urine, anaemia tests, as also ECGs, taking temperature, taking blood samples, checking sugar levels, pulse, etc. It is a fully equipped, state-of-the-art unit with medical equipment powered by cutting edge technology.
- The business is a private limited company with recognition from Indian Government. AmbuPod started about four years ago and till date around Rs. 48–50 lakhs have been spent.
- This year, the owners plan to come up with a more robust model of AmbuPod.
- So far, they have got over two million hits on social media, fifteen thousand likes and around three thousand comments and congratulatory messages.
- They received an award from Amrita Technology Business Incubators at their pitch fest in 2017 as Best Social Sector Impact Start-up.
- They have received a lot of leads from Africa. Recently, they signed an agreement with an African company, and have now become the official distributor.
- So far, they have planned a pilot project of twenty-five villages for which they have shortlisted eighty-six R.M.s.
- One of the biggest advantages that AmbuPod will have after implementing the project is that big data will help insurance companies, government and corporate understand which disease will hit which area at what time and what antibiotics to use during which time of the year. This helps streamline the stocking of medicines. It will thus support predictive analytics.
- AmbuPod is cost-effective. One AmbuPod that is fully equipped and is telemedically enabled (i.e. has a mobile clinic) costs Rs. five lakhs. The basic model costs Rs. 3.8 lakhs. An ordinary fully equipped basic life support ambulance costs around Rs. 8–10 lakhs, and the advanced ambulances cost around Rs. 15 lakhs.

- The owners have secured two patents and have now applied for an international patent.

5.5 Challenges Facing the Enterprise

AmbuPod was officially inaugurated in January 2017. But it is not yet earning revenue. The implementation of this business is still pending. The plan of the owners is to enter 50,000 villages out of the six lakhs villages in India in the first five years in any state that is willing to consider their different business models.

6 Conclusions and Discussion

The product has the potential of changing the face of rural health care. It is the future 'Uber' of health care. The fact that it will facilitate predictive analytics means that it will enable pre-emptive treatment/health care. It will help us decide how best to carry forward preventive care and prevent epidemics. As per the owners' research, around 5000 children die every day in India directly or indirectly because of lack of or access to health care. The service if implemented will reduce the large number of children dying on a daily basis.

7 Recommendations

Government of India should help the potential social entrepreneurs not only by providing information, advice and suggestion but also by linking the social entrepreneurs with the local and global funding agencies which provide finances for growth and development of social enterprises. It should coordinate and monitor the activities of social enterprises, and provide necessary support for their development. It should also arrange for financial rewards and incentives for good performance of the social enterprises.

The researchers also recommend that corporate should actively go in for such low-cost yet very useful healthcare services as provided by AmbuPod for fulfilling their CSR goals as per law. It will also result in meaningful and inclusive healthcare services being made available to poor and rural folks. NGOs also should consider cost-effective options like AmbuPod to facilitate their social services in the healthcare sector.

When we consider the model of entrepreneurship stipulated by A. Singh, we realize that Lynkambupod Pvt. Ltd. actually incorporates all the components. Dr. Lavanian recognized the opportunity for value creation. He had the capability of social entrepreneurs with which he mobilized resources along with a lot of innovation

to suit the given context of poverty and inaccessibility to some villages and created AmbuPod along with an operations strategy which would even provide employment and training to some persons.

There is a fundamental difference in the concept of scale between the business-corporate and social enterprise sectors. The success of business corporations is invariably measured by the size they attain: their incomes, profits, stock valuations and the wealth accumulated by their founders. Whereas the success of social enterprises must be measured by the impact they have on the lives of others. Indeed, the larger its impact and the smaller the organization itself is, the more successful it is. The main purpose of any social entrepreneurship is to bring about social change by changing the day-to-day realities at the local level. The local context thus shapes opportunities for social entrepreneurship and determines the strategies used. These strategies do reflect the entrepreneurial spirit in terms of resourcefulness and the ability to recombine the existing resources into new value creating configurations, and also a novel way of doing things.

AmbuPod has the potential of making such an impact if given the right impetus. It is a low cost way of making health care an inclusive one in India. The founder of the enterprise found the opportunity in the social context where he was working and decided to change the day to reality for the rural and poor. The fact that he was trying to bring quality health care to such vast segments of society made him choose strategies and recombine available resources into a wholly new concept of healthcare. He has also strategized to ensure that this value creating configuration in the form of AmbuPod can be implemented in a novel manner that will also provide employment. It is now up to the government, corporate and NGOs as also the village mukhiyahs to ensure that this benefits the vast majority of Indian rural and poor people first.

References

1. Santos, P.M.: A positive theory of social entrepreneurship. *J. Bus. Ethics* **111**, 335–351 (2012)
2. Trivedi, C., Stokols, D.: Social enterprises and corporate enterprises: fundamental differences and defining features. *J. Entrep.* **20**(1), 1–32 (2011)
3. Census of India: 2011, accessed from censusindia.gov.in on 31 Oct (2011)
4. Asian Development Bank: Key Indicators for Asia and the Pacific 2011
5. UNDP: Human Development Report 2011. Sustainability and Equity: A Better Future for All. UNDP, New York (2011)
6. Aggarwal, A.: India's services sector: gateway to development? *Econ. Polit. Wkly.* **XLVII**(26 & 27):119–123 (2012)
7. Shukla, M.: Reaching social entrepreneurs in India's Hinterland (2012, 18-November). Retrieved 2013, 15-July from Inspired-pragmatism Blogspot. <http://inspired-pragmatism.blogspot.in/>
8. Meher, R.: Livelihood, poverty and morbidity: a study on health and socio-economic status of the tribal population in Orissa. *J Health Manag.* **9**(3), 343–367 (2007)
9. WHO: World health Statistics 2013: A Wealth of Information on Global Public Health (2013)
10. Short, J.C., Moss, T.W., Lumpkin, G.T.: Research in social entrepreneurship: past contributions and future opportunities. *Strateg. Entrep. J.* **3**, 161–194 (2009)

11. Yujuico, E.: Connecting the dots in social entrepreneurship through the capabilities approach. *Socio-Econ. Rev.* **6**, 493–513 (2008)
12. Elkington, J., Hartigan, P.: *The Power of Unreasonable People: How Social Entrepreneurs Create Markets that Change the World*. Harvard Business Press, Boston, MA (2008)
13. Khanna, T.: India's Entrepreneurial Advantage. *McKinsey Quarterly*, 2004 Special Edition, p. 111–114 (2010)
14. Intellecap: On the Path to Sustainability and Scale: A Study of India's Social Enterprise Landscape. Intellecap, Hyderabad (2012)
15. Sharma, A.: India's welfare gamble: add 100 million to the Dole. *The Wall Street J.* June 8, pp. A1, A16 (2010)
16. Korsgaard, S., Anderson, A.R.: Enacting entrepreneurship as social value creation. *Int. Small Bus. J.* **29**(2), 135–151 (2011)
17. Singh, A.: *The Process of Social Value Creation*. Springer, India, p. 52 (2016)
18. Polonsky, M.J., Grau, S.L.: Evaluating the social value of charitable organisations: a conceptual foundation. *J Macromark* **28**(2), 130–140 (2008)
19. Corner, P.D., Ho, M.: How opportunities develop in social entrepreneurship. *Entrep. Theory Pract.* **34**(4), 635–659 (2010)
20. Mizik, N., Jacobson, R.: Trading off between value creation and value appropriation: the financial implications of shifts in strategic emphasis. *J. Mark.* **67**(January), 63–76 (2003)
21. Beyond Profit (Intellecap): Indian Social Enterprise Landscape Survey. Data self-reported by SEs (2010)

Quality Assurance Practices and Techniques Used by QA Professional in Continuous Delivery



Anish Cherian , Raju Ramakrishna Gondkar and S. Suresh Babu

Abstract In traditional software development life cycle like waterfall and V-model, there are staged gates, where at the end of each phase, there are strict quality checks being done. If the check passes, the team is allowed to go to the next phase. After the emergence of agile software development, continuous integration and continuous delivery are widely practiced in most of the teams. Quality assurance checks happen as soon as the developer checks in and the code is further built and tested in the build pipeline. In continuous delivery, quality checks are being done instantly. The responsibility of the software quality assurance (SQA) professional is to ensure that the right process is defined in the project and the process works fine. The SQA conducts planning, reviews, and audits to ensure this. In the traditional SDLC like waterfall and V-model, audits and reviews are conducted before the end of each phase. There are lots of reference standards for this. However, the way software quality assurance should be done is very different in continuous delivery mode of development. There is less literature and standards available about how an SQA professional should operate when team is using continuous integration and continuous delivery. This paper provides an overall framework of continuous delivery quality assurance model, and the implementation results in information and communication technology (ICT)-Internet of things (IOT) domain.

Keywords Quality assurance · Continuous delivery · DevOps · SQA · DevOps

A. Cherian · R. R. Gondkar
CMR University, Bangalore 560043, Karnataka, India
e-mail: anish.15phd@cmr.edu.in

R. R. Gondkar
e-mail: raju.r@cmr.edu.in

S. S. Babu
Huawei Technologies India Private Limited, Bangalore, Karnataka, India
e-mail: sureshbabu.s@huawei.com

1 Introduction

Continuous Delivery a software development approach where as soon as the developer checks-in the code, the build is run, goes through a deployment pipeline and the completed build is ready to be delivered to the customer. The quality assurance checkpoints are built into the deployment pipeline. The key objective of this paper is to provide continuous delivery quality assurance practices to be followed by a quality assurance professional (also called SQA) in information and communication technology (ICT) domain. Also, the implementation results will be depicted after deployment of the practices in a particular project.

2 Comparative Analysis of Traditional Versus Agile Quality Assurance

Software development life cycle is a systematic way of software development. There are various lifecycle models like waterfall, V process model, spiral, rapid application development, agile, lean, and others. Table 1 contains the quality assurance activities in traditional and agile software development life cycle.

In waterfall model, quality assurance activities are carried out at each phase through staged quality gates and supporting practices. Once requirements are defined, it is validated and verified with the help of reviews, prototyping, and model validation. Code reviews are conducted through formal inspection mechanism. Testing and release are done after the verification and validation mechanism. In agile method,

Table 1 Comparative analysis of traditional versus agile quality assurance

Activities	Details	Traditional QA practices	Agile QA practices
Requirement analysis	Requirement understanding, system feasibility and software plan	Requirements V&V—requirement reviews	Daily standup meetings, product backlog review
Design	System and software design	Design V&V—model checking, checklists	Simple design collaborative review
Coding	Implementation and unit testing	Inspection, reviews	Pair programming, refactoring, unit testing
Test	Integration and system testing	Integration and acceptance testing	Continuous integration, onsite customer
Release	Release and maintenance	Tools, release flow approval, closure	Sprint review, retrospection

quality assurance activities are built into the activities at iteration level through practices like pair programming, test-driven development (TDD), automated testing, refactoring, continuous integration/delivery, and others. Agile team uses continuous delivery approach where as soon as the code is checked in, and the build runs in deployment pipeline (divided into stages) in order to give quick feedback about the quality of the check-in. The quality checkpoints are built into the deployment pipeline. DevOps is also a relevant practice where the development and operations team work together closely to reduce the time between committing a change to a system and the change being placed into normal production, while ensuring high quality.

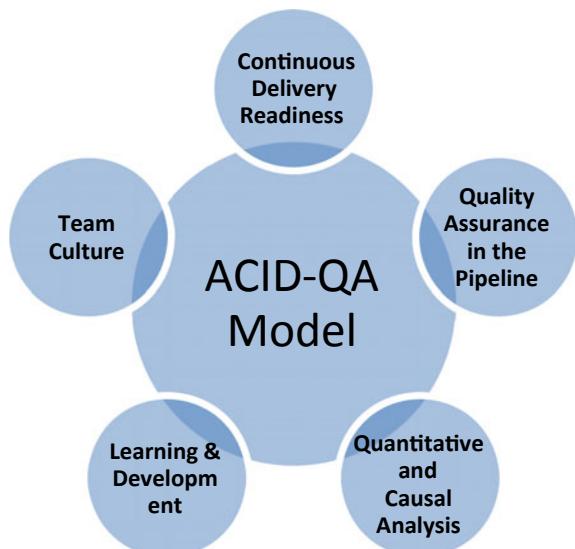
3 Proposed Model

In continuous delivery mode of development, quality assurance should be viewed from the following perspectives (Fig. 1).

Continuous Delivery Readiness

Requirement documentation at story level should be at a granularity that a story can be developed in 2–3 days. This can help the developer to check in the story at frequent intervals. This will also help in iterative delivery practices. System architecture should consider deployability, modifiability, monitorability factors. CI architecture should be based on the system architecture. This helps in having independent CI system based on the component level dependency, and there can be faster feedback provided. Build pipeline should be based on the CI architecture. Environment

Fig. 1 Agile continuous integration and delivery—quality assurance (ACID-QA) model



provisioning should be done based on customer requirement analysis (customer environment inbuilt into the pipeline).

Quality Assurance in the Pipeline

In the deployment pipeline, inspection tools like static code quality tools (e.g., FindBugs), duplicate code and dead code (e.g., Simian), architecture and design monitoring (Structure 101, JDepend tools) can be integrated. It is important to ensure that right rules are configured for these inspection tools. While running the unit testing and functional testing, coverage tools can be integrated at unit and functional test level (functional/line/branch coverage). Since test code is a safety net for the system, it is equally important to ensure the quality of the test code. Static quality check tools like FindBugs can also be integrated for ensuring quality of the test code. In the continuous deployment pipeline, test suite should have the right composition of unit, function, UI, non-functional test cases. Effective test pyramid practice can be followed by teams. In order to ensure security assurance, tools for scanning like Nessus, NMAP, and security test cases can be integrated into the build pipeline. Threat modeling technique can be used for generating the test cases. Non-functional test cases related to performance, memory leak, maintainability, etc., can also be integrated into the pipeline. The infrastructure of the deployment pipeline can also be tested. Build script quality should be ensured through practices like abstraction, modularization, and coding guidelines adherence.

Quantitative and Causal Analysis

Build results can be quantitatively analyzed for the overall quality improvement. Periodic assessments and causal analysis can be done for these build results, and it can bring systematic improvements like refactoring flaky test cases, quality improvement, and technical debt improvement for defective modules.

As shown in Table 2, the version build is continuously passing, but the functional build is failing. There can be detailed analysis further and based on the test cases, or issues related to static checks can be corrected.

Learning, Development, and Team Culture

Team culture is the foundation for effective continuous delivery. There should be strong cross-cutting collaboration between development, test, security, operations, and related roles. Knowledge can be cross-pollinated through knowledge management practices. Team members can be trained on continuous delivery mode

Table 2 Analysis of the build pipeline

Build	Private build	Version build	Function build	“ties” build	Deployment build
Build 01	Pass	Pass	Fail	Fail	Fail
Build 02	Pass	Pass	Pass	Fail	Fail
Build 03	Pass	Pass	Fail	Fail	Fail
Build 04	Pass	Pass	Fail	Fail	Fail
Build 05	Pass	Pass	Fail	Fail	Fail

of requirement development and analysis, system architecture → CI architecture → build pipeline, coding, testing, build scripting and related practices.

Implementation of the Model in the Internet of Things (IOT) Domain

The quality assurance model is further piloted in IOT protocol platform. The practices from the model were selected and implemented in the latest version of the platform. The key requirements in this version were related to IOT—lightweight Internet protocol. The Internet of things (IOT) comprises f network of all devices which is embedded inside enabling these devices to connect and exchange data. Lightweight IP (lwIP) is an open-source TCP/IP implementation which is designed for embedded systems. The focus on lwIP is to effectively use the resources and still use the TCP stack. This makes lwIP suitable for embedded systems. Lightweight IP uses following protocols in each of the layers: Internet layer: IP, ICMP, IGMP, ICMPV6; transport layer: UDP, TCP; application layer: DNS, SNMP, DHCP; link layer: PPP, ARP, Neighbor Discovery Protocol for IPV6.

3.1 Trade-off Analysis

The motivation behind implementation of the proposed model in this project is to improve the quality of the non-conformances found during the QA audit. In order to select the right practices from the model, trade-off analysis was conducted. The major parameters considered for the trade-off analysis were: 1. Deployment Effort: Effort spent by the SQA professional and the team should not be very high for the deployment of the selected practices. 2. Competency: Time required to improve the technical skills of the SQA professional should be relatively less. 3. Voice of Stakeholders: Feedback from the stakeholders of the project in a scale of 1–5. 4. Ease of Process Change: It should be easy to do the process change in the project.

Let us say, “increase in the number of product quality issues identified during QA audit” is considered as (Y). Based on the proposed model, following aspects are considered as critical to quality (CTQ) for this. We name them as following: y_1 = Continuous Delivery Readiness, y_2 = Quality Assurance in the Pipeline, y_3 = Quantitative and Causal Analysis, y_4 = Learning & Development, y_5 = Team Culture. For piloting the model in telecommunication protocol project, CTQ is selected based on the below trade-off analysis (Table 3).

A scoring was given in a scale of 1–5 for the above parameters. The Y_n which scored more than 15 was considered for piloting in the project. Pilot was planned majorly focusing Y_1 —continuous delivery readiness and Y_2 —quality assurance in the pipeline. For each CTQ identified which are the output of the process, input of the process ($X_1, X_2\dots$) is identified. Below listed are the input of the process ($X_1, X_2\dots$) for each CTQ identified:

$y_1 = f(X_1, X_2)$, $y_1 = \text{Continuous Delivery Readiness}$ where X_1 = Test Suite Organization issues, X_2 = Environment Provisioning based on customer requirement analysis (Customer environment inbuilt into the pipeline).

Table 3 Trade-off analysis

Weighted result	Model deployment effort	Competency	VOC	Process change	Total
y_1 —continuous delivery readiness	5	4	4	4	17
y_2 —quality assurance in the pipeline	4	5	5	5	19
y_3 —quantitative and causal analysis	2	4	2	4	12
y_4 —learning & development	3	3	3	3	12
y_5 —team culture	2	3	3	3	11

$Y_2 = f(X_1, X_2, X_3)$, **Y₂ = Quality Assurance in the Pipeline** where X_1 = Inspection/Static Quality Assurance at Feature Level, X_2 = Rules for Rules, X_3 = Security Assurance.

Measuring the Pilot Improvement using Statistical Approach

The proposed model is piloted in the IOT platform project. Four cycles of audits are conducted based on the learning's from the model.

3.2 Key Audit Findings

See Tables 4 and 5.

Audit Findings Summary

Table 6 is summary of all the issues identified as part of four cycles of audit and compared the number of findings before the model was piloted.

3.3 Test of Hypothesis

Statistical test is conducted to verify whether the model has brought the improvements. Test of hypothesis is done to check improvement of a model about the CTQ identified. Normality test: Both data of product quality issues found in the year 2017, and in Q1, 2018 is normal as p -value is > 0.05 (Table 7).

Table 4 Key audit findings from the audit based on the model piloted

Issue category	Audit before the model pilot	Audit after the model pilot and issues observed
Right rules are not used in static tool	SQA assumes that right rules are configured for static tools in the CI system. SQA sometimes do not possess deeper technical competency to check the rules configuration	Technical skill to understand the rule file is developed by SQA. In the audit, it is found that around 10 rules were masked in the static rules file. Out of 10 rules masked, there is no analysis available from team for 6 rules which are masked. Team further checked on why these rules are masked and took further actions. SQA was able to find these issues independently. We developed a plug-in (which runs in CI system), which will automatically check the version of the static tool plug-in used, to ensure always the latest version is used and also provide list of rules which are masked
Feature-level technical debt using inspection quality assurance	Total technical debt of the project is given. There is no detailed analyses on which are the risky modules	SQA did deeper analysis of the static tool issues and prepared a map at module level. As an audit output, provided module-wise technical debt using static analysis and highlighted the top risky modules in the project. (Refer Table 3 for details.)
Test suite not organized at feature level	Test suite organization was checked by SQA	Test suite organization is audited by SQA. As an audit output, it is found that test suites are not organized at feature level which will have cascading impact in identifying the feature stability based on test case failures. Test case naming conventions are found to be wrong which impacts the identification of the feature stability and running the test suite in the deployment pipeline effectively

(continued)

Table 4 (continued)

Issue category	Audit before the model pilot	Audit after the model pilot and issues observed
Not all supported customer environment is built into the pipeline	CI environment was never checked by SQA	Audit is done to check the continuous delivery readiness, where it is observed that CI is always run only in Windows environment and in 1 non-Windows environment, though the platform supports several other environments, including SUSE Linux, Red Hat Linux, VxWorks, different versions, and no clear strategy on choosing the environments in build pipeline. All the test cases are run only in Windows environment

Table 5 Sample feature-level technical debt

Feature list	Code size (KLOC)	Coding style—P-CLint	Unsafe functions, security	Code duplication (LOC)	Max code complexity
XXX	6.9	0	2	0	<10
XXX	1.23	0	0	0	<10
XXX	0.99	0	76	0	23
XXX	2.24	0	0	0	<10
XXX	9.12	0	3	35	<10

Table 6 Audit findings' summary

Year 2017 issue list	Year 2017	H1 2018	H1 2018 Issue List
QA Audit 1	5	13	Q1-January 2018 CI audit
QA Audit 2	7	14	Q1-February 2018 CI audit
QA Audit 3	9	16	Q1-March 2018 CI audit
QA Audit 4	9	18	Q1-March 2018 CI audit

Table 7 Test of hypothesis result

Type	Description																																			
Null hypothesis (H0)	Number of product quality issues identification has not improved. The mean difference between paired observations is zero																																			
Alternate hypothesis (H1)	Number of product quality issues identification has improved. The mean difference between paired observations is not zero																																			
Test being used	Paired T-Test																																			
Test results	<p>Paired T-Test and CI: Year 2017, Q1 2018</p> <p><i>Paired T for Year 2017—Q1 2018</i></p> <table> <thead> <tr> <th></th> <th>N</th> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> </tr> </thead> <tbody> <tr> <td>Year 2017</td> <td>4</td> <td>7.50</td> <td>1.91</td> <td>0.96</td> </tr> <tr> <td>Q1 2018</td> <td>4</td> <td>15.25</td> <td>2.22</td> <td>1.11</td> </tr> <tr> <td>Difference</td> <td>4</td> <td>-7.750</td> <td>0.957</td> <td>0.479</td> </tr> <tr> <td>95% CI for mean difference:</td> <td>(-9.273, -6.227)</td> <td></td> <td></td> <td></td> </tr> <tr> <td>T-Test of mean difference = 0 (vs ≠ 0):</td> <td>T-Value = -16.19</td> <td></td> <td></td> <td></td> </tr> <tr> <td>P-Value</td> <td>= 0.001</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>		N	Mean	StDev	SE Mean	Year 2017	4	7.50	1.91	0.96	Q1 2018	4	15.25	2.22	1.11	Difference	4	-7.750	0.957	0.479	95% CI for mean difference:	(-9.273, -6.227)				T-Test of mean difference = 0 (vs ≠ 0):	T-Value = -16.19				P-Value	= 0.001			
	N	Mean	StDev	SE Mean																																
Year 2017	4	7.50	1.91	0.96																																
Q1 2018	4	15.25	2.22	1.11																																
Difference	4	-7.750	0.957	0.479																																
95% CI for mean difference:	(-9.273, -6.227)																																			
T-Test of mean difference = 0 (vs ≠ 0):	T-Value = -16.19																																			
P-Value	= 0.001																																			
Inference	Ho is false (number of product quality issues identification has improved). The test gives a <i>p</i> -value of 0.001 which is < 0.05. Thus, we can reject the null hypothesis at the 0.05. α -level concludes that the number of product quality issues identification has improved after the actions taken																																			
Results	Number of product quality issues identification has improved																																			

4 Conclusion and Future Work

Based on our literature review, we have found that most times, quality assurance is mixed with testing. While there is nothing wrong in considering testing as quality assurance, it needs to be kept in mind that quality assurance is to do with more holistic approach or activities for quality. We have analyzed the key concepts from the perspective of quality assurance and then mapped to the general role of a software quality assurance and then categorized the key concepts in the above framework. Based on the pilot of the continuous delivery quality assurance framework in four different audits, it is evident that the QA is able to identify more product quality-related issues based on this framework. In this approach, SQA professional needs to have deeper understanding of the technical areas of continuous delivery, inspection/static quality assurance, and other areas related to DevOps. This will help in finding the product quality-related issues in systematic manner using this model. This model gives room for more options to automate the quality assurance parameters and bring predictability in the product quality. The initial implementation results of the proposed QA audit framework (in continuous delivery mode of software development) have been encouraging. There are key actions planned to automate some of the quality assurance checkpoints, so the product quality issues can be found automatically.

Acknowledgements I thank my employer and my university professors for the support provided, without which it would not have been possible to complete this research paper. The article represents only my own views and does not necessarily reflect those of my employer.

References

1. Bass, L., et al.: DevOps A Software Architect's Perspective, 1st edn. Addison-Wesley, New York, NY (2015)
2. Chen, L.: Towards architecting for continuous delivery. In: 12th Working IEEE/IFIP Conference on Software Architecture, IEEE, Canada (2015)
3. Jez Humble On Continuous Delivery Homepage: <http://continuousdelivery.com/>

Sensponics: IoT-Enabled Automated Smart Irrigation and Soil Composition Monitoring System



Prasun Guchhait, Pranav Sehgal and Vidyadhar J. Aski

Abstract ‘Water’ is one of the key components available to mankind. All living things consist mostly of water; e.g., the human body is made up of 67% of water. Water is the crucial component of life and is essential for sustenance, so as it is a most vital component for upgrading agricultural productivity, and hence, the utilization of water in a most efficient manner is the key concept we must follow to improve the farming/gardening in the nation. Sensponics helps the farmers/gardeners to distribute water to crops/plants by providing them with water when they need the water, and this helps to prevent wastage of water and soil degradation. In this project, we will develop an automated smart monitoring and irrigation system that helps farmers/gardeners to know the status of their crops/plants from home or from any part of the world. This system helps farmer/gardeners to irrigate the land in a very organized manner based on soil humidity, atmospheric temperature and humidity, and water consumption of the plant. Water surplus irrigation reduces plant production and degrades soil fertility and stimuli ecological hazards like water wasting and soil degradation. The smart system not only provides comfort but also reduces energy conservation, efficiency, and time-saving. Nowadays, farmers are not financially stable to use industry graded automation and control machine which are high in cost. So, in this project, we will implement a concept of Internet of things (IoT) to read the data from sensors using Arduino Uno and send it to ThingSpeak, an open-source cloud to store and analyze the data of sensors.

Keywords Water management · Soil degradation · Internet of things (IoT) · Agricultural productivity

P. Guchhait (✉) · P. Sehgal · V. J. Aski

Department of Computer and Communication Engineering, School of Computing and Information Technology, Manipal University Jaipur, Jaipur 303007, Rajasthan, India
e-mail: prasunguchhait1997@gmail.com

P. Sehgal
e-mail: pranavsehgal12@gmail.com

V. J. Aski
e-mail: vidyadharjinnappa.aski@jaipur.manipal.edu

1 Introduction

The significance of building an automation system for agricultural fields, homes, offices increases day by day. Automation in irrigation makes an efficient use of the water, electricity and reduces much of the wastage. Aeroponics and hydroponics are some of the existing methodologies that help in efficient irrigation by using air and water, respectively, as a medium to provide nutrients to the plants. Sensponics () is a proposed methodology which makes an efficient use of water for automation in irrigation. The crux of Sensponics is an Arduino Uno microcontroller. The advantages of the system are in threefolds.

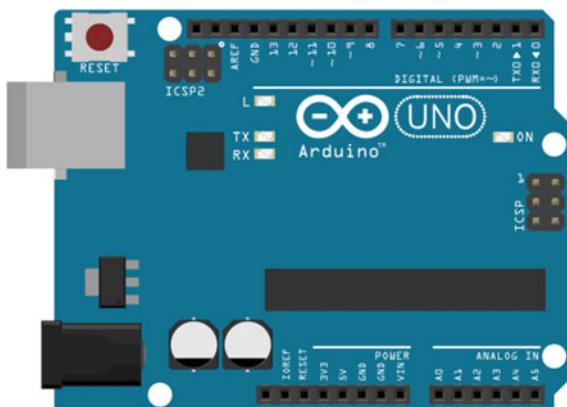
- Conservation of resources for proper utilization.
- Farmers would be able to manage the water requirements by automating farm or nursery irrigation.
- Avoided unplanned irrigation and reduced overwatering will enhance plant's growth.

Arduino Uno traits with an open-source hardware board designed on an 8-bit Atmel Automatic Voltage Regulator microcontroller. Current models consist of a USB interface, six analogue input ports and 14 digital I/O ports that allow the user to attach the various extension to the microcontroller (Fig. 1).

2 Literature Review

After extensive research in the agricultural field, many researchers found that the agricultural productivity is decreasing day by day. Many researchers have proposed some suitable methodologies to overcome the problem of agricultural productivity.

Fig. 1 Arduino Uno
ATmega328P



Chandan Kumar, Sahu et al. proposed a system on “A Low-Cost Smart Irrigation Control System.” It consists of wireless sensors. The heart of the system is Raspberry Pi, and it is used to send data like images and text messages through Internet communication to the microcontroller processes [1].

K S. Nemali et al. proposed irrigation systems which are automated using dielectric moisture sensor information on the content of the soil. It is used to control actuators to save water [2].

Kapoor, Ayush, et al. proposed a system on “Implementation of Internet of Things (IoT) and image processing in smart agriculture.” It combines image processing techniques and the Internet of things to smartly monitor the growth of plant [3].

With the use of various proposed techniques and introduced technologies, we can increase the productivity of fields and reduce the manual effort and resources to be required.

3 Proposed System

The proposed system as shown in Fig. 2 consists of diverse types of sensors such as temperature and humidity sensor, soil moisture sensor, water-level sensor light sensor, and various actuators like LED, water submersible connected to Arduino to perform various predefined tasks one by one and send the relevant data to the cloud.

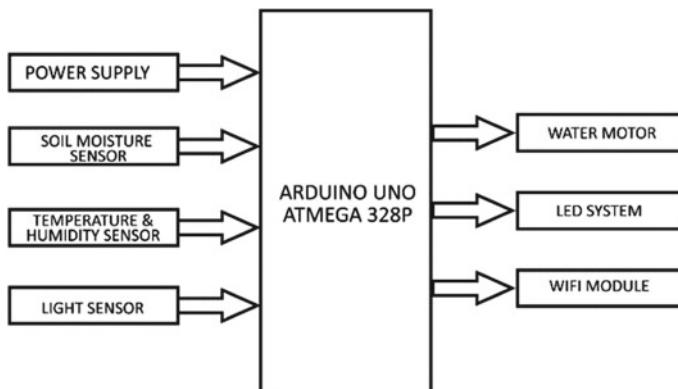


Fig. 2 Proposed system: Sensponics

3.1 Sensor and Actuator Used

The proposed system consists of different sensors and actuators which were required to do specific tasks. The individual sensors and actuators are listed in Tables 1 and 2 with their descriptions.

3.2 Flow Control of the System

Since different sensors and actuators were there in the proposed system, there should be a proper flow of controls to each module so that the system can work efficiently and effectively. Figure 3 shows the flow of control of the system.

3.3 Cloud Storage

The data collected from various sensors need to be stored somewhere either offline or online; storing data offline can result in a big disadvantage for the system as storage space (memory) is required; the data cannot be fetched over the Internet, so we use a cloud storage solution named [thingspeak.com](#), and the advantages of [thingspeak.com](#) are in threefolds.

Table 1 Sensors used in this system

Name	Description
Soil moisture	An electrode-based sensor which measures the current flow between the medium where the electrodes were placed
DHT 11	Capacitive humidity sensor to measure the atmospheric humidity and temperature
LDR	The diode-based sensor produces an output signal that indicates the intensity of light by reading the radiant energy
Water level	The water-level sensor senses the liquid water droplet depth to measure the water level in the water tank

Table 2 Actuators used in this system

Name	Description
Wi-Fi module	The ESP8266 Wi-Fi module is a low-cost microchip with self-contained SOC capable to access to Wi-Fi network
Water submersible	The water submersible is used to supply water to the plant, and it is a 9 V motor powered by a 9 V battery
Relay	The 5 V relay is an electrical switch used to control water submersible using electrical signals

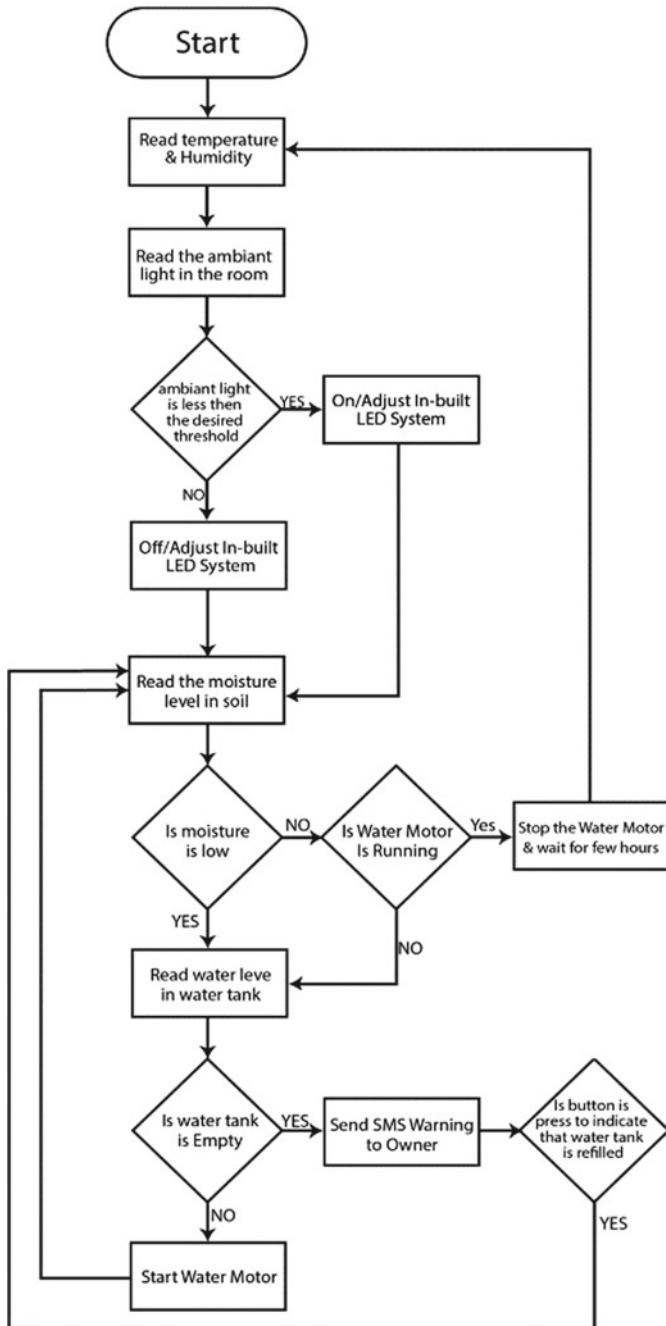


Fig. 3 Flowchart of Sensponics

- The cloud is a free and open source.
- Data storage and data fetching are very convenient for using API calls.
- The data stored in the cloud displayed in a graphical format which helps in better understanding of the data.

To communicate between different devices and interfaces, there is a requirement of a common file format supported by both the interfaces. ThingSpeak cloud supports different types of file format but in this system, all the data exchange is done in JSON file format; only Fig. 4 shows the communication of Sensoponics with ThingSpeak cloud.

3.4 Event-Triggered Notification

The proposed system can read sensory data and store them in cloud, but to make it more effective, there must be some notification system which notifies the user in event-driven bases, and to achieve this, we have to create a local Web server which fetches data from cloud and store them in database and also trigger the message service in event-driven bases. Sensoponics will use way2SMS messaging service to send the message to the user. The advantages of way2SMS are in twofolds:

- The messaging service is free.
- Python libraries are available to send SMS from scripts.

creating a local server in benefits for future purposes because these data sets can be used in data analytics or in other frameworks to understand the growth of plants.

4 Methodology of Sensoponics

Sensoponics is an IoT-based system that not only helps the plants in irrigation but also helps in monitoring the needs of the plants. The sensors in Sensoponics read different parameters of a plant, i.e., temperature, humidity, light consumption, water

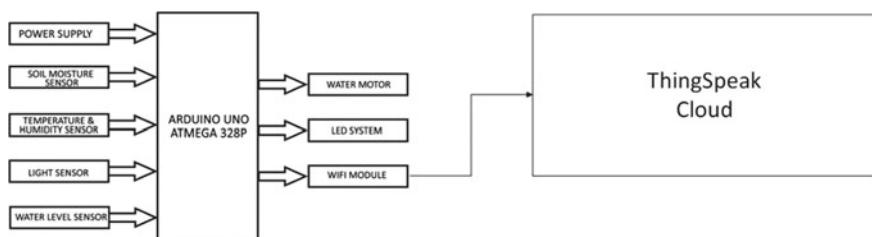


Fig. 4 Communication between Sensoponics and [ThingSpeak.com](https://thingspeak.com)

consumption, and send the relevant data to the cloud; Fig. 5 shows how the sensors connect to Arduino.

5 Experimental Results

After modeling and developing a prototype model (Fig. 6), it has been tested with a real plant and connected to the cloud to store sensor data.

Similarly, the data uploaded to the cloud are shown in Figs. 7 and 8, ThingSpeak supports REST API and MQTT API to upload data to the cloud, but in this system, only REST API is used using GET method to upload the data to the cloud.

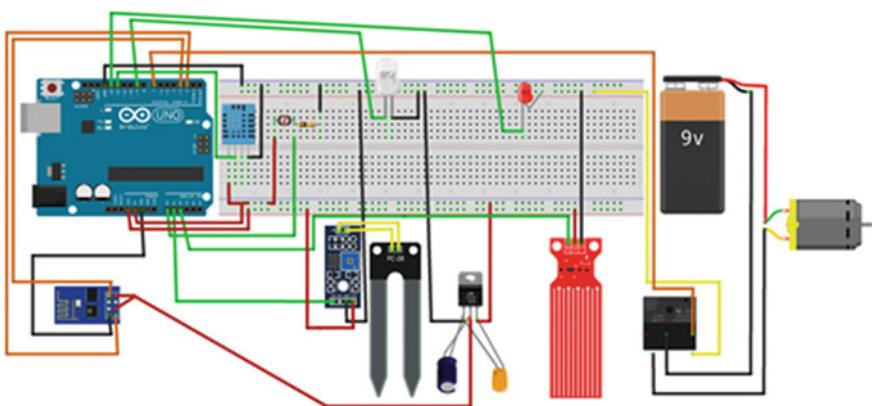


Fig. 5 Sensoponics circuit layout



Fig. 6 Sensoponics: prototype model

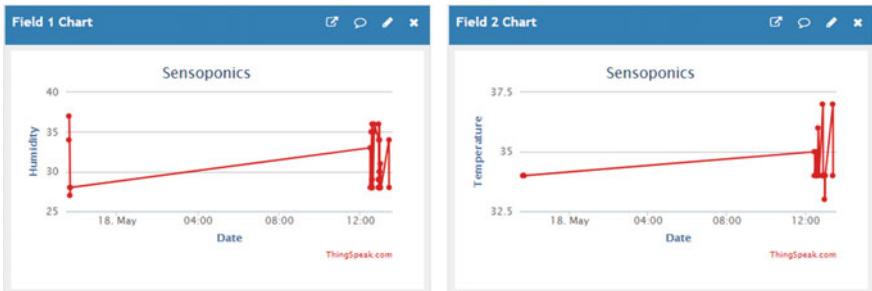


Fig. 7 Humidity and temperature data uploaded in the cloud

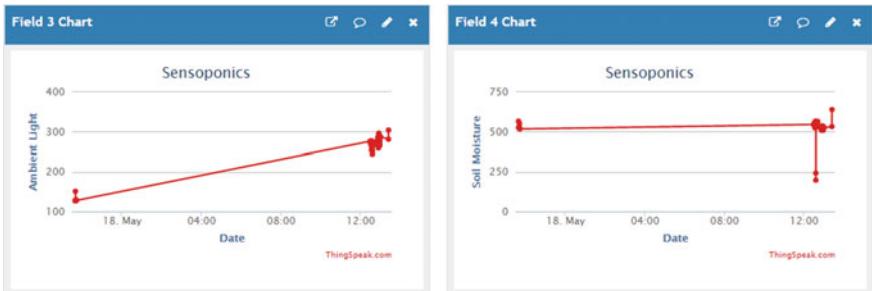


Fig. 8 Ambient light and soil moisture data uploaded in the cloud

6 Conclusion

Farmers, gardeners, or people who want to grow fruits and vegetable in the house have to use old traditional techniques to grow the plants in their home which may sometimes fail due to various natural or artificial conditions and calamities; the old traditional techniques tend to use much amount of water and also are responsible for soil infertility, water salinity, and even wastage of water; moreover, areas with less rainfall like some places in Rajasthan are difficult to irrigate plants and crops without consuming many resources. Hence, there is a requirement of an automated system which tends to help in precise monitoring and control the water requirements of the plant. Installing Sensoponics in vicinity ensures the efficient use of water, time, and resources; moreover, the system uses Arduino Uno which ensures low electricity consumption and reliability of growing plants perfectly.

References

1. Sahu, C.K., Behera, P.: A low cost smart irrigation control system. In: 2015 2nd International Conference on Electronics and Communication Systems (ICECS). IEEE, 2015
2. Nemali, K.S., van Iersel, M.W.: An automated system for controlling drought stress and irrigation in potted plants. *Scientia Horticulturae* **110**(3), 292–297 (2006)
3. Kapoor, A., et al.: Implementation of IoT (Internet of things) and Image processing in smart agriculture. In: International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS). IEEE, 2016

A Novel Trust Scheme in Semantic Web



Suresh Kumar and Neera Chaudhary

Abstract In this era of information technology, trust and proof are the most prominent research areas of the Semantic Web (SW). Many researchers working in SW have realized the importance of trust and proof for the protection of data available on the Web. Broadly, there are two trust approaches available in the literature. One is policy-based approach, and second is reputation-based approach. In this research paper, we have proposed a mathematical novel trust scheme in SW based on reputation approach for selecting an education institute for getting admission in particular course in SW. We have identified ten trust parameters for selecting an institute for getting admission like placement, faculty, accreditation, extracurricular activities, transport facilities, hostel facilities, health facilities, basic infrastructure, fee, and attendance. Based on these parameters, we have designed a questionnaire. Then, we have made a survey on 200 engineering students by distributing questionnaire among them. We have presented a sample of questionnaire in this paper. We have also presented the results of our scheme in the form of pie and bar graphs. Our results validated our mathematical approach.

Keywords Semantic Web · Educational institutes · Trust schemes · Trust policies

1 Introduction

Trust can be achieved by creating what is known as a “Web of Trust” [1]. Trust is based on experience, and by tracking and propagating trust among Web sources in a similar fashion to the way trust is created and maintained in a human community [1]. In this world of the Internet, Semantic Web (SW) is considered as Web

S. Kumar (✉) · N. Chaudhary

Ambedkar Institute of Advance Communication Technologies & Research Government
of NCT of Delhi, New Delhi, India
e-mail: sureshpoonia@aiactr.ac.in

N. Chaudhary
e-mail: neerachaudhar18@gmail.com

of today. Tim Berners-Lee presented the SW stack in year 2000. Trust is on the top of the stack, and it is considered as an essential element because it provides security to all the layers which are below it. In this paper, initially we have presented all the existing trust schemes which compute trust in the SW networks, e.g., peer-to-peer and social networks, and then, we have defined and validated our mathematical approach. Social network is a network in which persons relate to each other (shows personal relationship). It shows the interaction between persons. It is a medium through which users can communicate with each other by sending messages, e-mails, posting pictures, images, videos, writing comments, and so on [2]. On the other hand, peer-to-peer network is a network in which peers refer to the computer systems that are interlinked and connected together through the Internet. There exists a client–server relationship between peer-to-peer systems. Each computer on the network is either a client or a server. We have found in the literature that there are broadly two trust approaches in SW, i.e., policy-based and reputation-based [3–5]. The policy-based approach is based on exchanging credential or personal information for communication between client and server to establish trust. Reputation-based approach is based on past experience, and rating is given by other users [3–5]. In this paper, we have proposed a reputation-based mathematical novel trust scheme in SW. In our proposed scheme, we focus on how a student could trust on a particular institute before getting admission in a course. We have identified ten basic parameters of preferences of students for selection of particular institute before admission. We termed these parameters as trust parameters. The parameters are placement, faculty, accreditation, extracurricular activities, transport facilities, hostel facilities, health facilities, basic infrastructure, fee, and attendance. Our objective is to design a trust scheme which aim was based on educational institute. We have done a survey on 200 engineering background students to know their preferences to select particular institute for getting admission in a particular course. Based on these parameters, we have designed a questionnaire. Then, we have made a survey on 200 engineering students by distributing questionnaire among them. We have also presented a sample of questionnaire in this paper. We have presented the results of our scheme in the form of pie and bar graphs. Our results validated our mathematical approach.

2 Trust Approaches in Semantic Web

There are broadly two main approaches of trust management in SW networks [6, 7], which are as follows:

Policy-Based Approach—This approach is based on exchanging credentials or personal information for communication between client and server to establish trust [3–5].

Reputation-Based Approach—In this approach, trust is established by past experience and rating is given by other users. In this approach, decision takes place based on recommendation given by other users [3–5].

3 Trust Mechanisms in Semantic Web

- **Context-Based Trust Mechanism**—In this mechanism, use of situational information has been taken that is the situation in which information has been assumed.
- **Content-Based Trust Mechanism**—In this mechanism, we make use of standard and dictum jointly with information content and also use the information that relates to the given topic and is provided by some other persons which are working on the same topic.

4 Goals for Trust and Policy Management

There are mainly two goals of trust and policy management which is defined as follows [8, 9]:

- Interoperability:—It is related to how to enable different autonomous communicating parties to bilateral agreements.
- Scalability:—Actually, there is no single authority and no single operational model of trust management available in the world. Scalable systems must deal with variation in national, corporation, and personal policies.

5 Trust Policies

Policies are some rules [10] set by the organization; they are the predefined standards. A trust policy [11] is a set of rules accepted to examine and calculate the trust for particular information related to a specific condition.

- One has to trust on professors in their respective research area.
- News shown on the media is sometimes manipulated and then shown, so one has to trust media only on sports news. We have to allow similar range of trust policies in SW.

6 Trust Applications

There are following trust applications [2, 11]:

- Trust bot
- Trust mail
- Marketing.

7 Related Work

Friend of a friend (FOAF) ontology is the basic for every trust approach or scheme. In every social network, FOAF is used to build a trust. A trust can also be encoded using other information. Data presented on the social network by the help of FOAF module/vocabulary is worthwhile information and widespread data in SW [12, 13].

Recommender systems are used to help the user to find out the items or information in which user is interested. To carry out recommendation, there are two ways [12]:

1. Calculate similarity between the items and the recommended items in which user is showing interest.
2. Calculate similarity between user and recommended items which are liked by similar users.

Second recommendation system is also known as collaborative filtering. The concept of collaborative filtering has been applied in many contexts to calculate trust. Institution trust is not the first attempt made for predictive recommendations about institute. There are recommendations made on films like film trust, recommends, film-conseil, and movie lens [12].

8 Technical Survey of Trust Schemes

We have surveyed in the literature regarding many trust scheme recommender system-based Web site. In one survey, we came across a film trust Web site [12]. By taking the example of that Web site, we are going to explain you how they computed trust by using mathematical concept. For computation, they considered a rating system. That rates the film based on the ratings that how we could trust on the Web site. In that work, they had used the concept of FOAF to trust in case of no direct connection. The formulae for computation of trust [12] are:

$$r_{sm} = \frac{\sum_{i \in s} t_{si} r_{im}}{\sum_{i \in s} t_{si}}$$

where r stands for rating, s for node m for movie, i is node and t for trust value. Then, in another survey we have found over a network in a network how one user could trust another. In that scenario, the trust was computed [2] as

$$t_{is} = \frac{\sum_{j=0}^n \begin{cases} (t_{js} * t_{ij}) & \text{if } t_{ij} \geq t_{js} \\ (t_{ij}^2) & \text{if } t_{ij} < t_{js} \end{cases}}{\sum_{j=0}^n t_{ij}}$$

where t is trust, and $i, j, \text{ and } s$ are nodes.

9 Proposed Trust Scheme

As we have seen trust on film Web site and trust on networks. We have decided to make a scheme to trust the educational institute Web site. We can compute the ratings mathematically. To compute the trust, we have taken the formula:

$$t_i = \frac{r_i}{\sum_{i=1}^{i=n} r_i} \quad (1)$$

where t is trust value and r is rating of particular institute i .

9.1 Survey Questionnaire

In our methodology, we have taken a sample of 200 students for which we have conducted the survey and collected the data. For this, we have prepared a survey questionnaire and given to students and ask them for different ratings by predefined rating scale. The rating scale is 1–5. From 1 to 5, the preferences are in descending order. The number 1 has the highest priority, whereas number 5 has the lowest priority. In the same way, the trust values are also rated from 0.1 to 0.5. The value 0.1 means the institute is highly trustable, and the value 0.5 refers to least trustful.

The sample of questions from questionnaire is mentioned below:

1. At the time of admission what rating (1–5) you give to select the institute to the following parameters?
 - a. Faculty
 - b. Infrastructure/laboratories
 - c. Placement
 - d. Extracurricular activities
 - e. Accreditation of the institute like NBA “A” Accreditation, AICTE, and UGC
2. Should there be a rigid attendance system to ensure students attend diligently?
 - a. Yes
 - b. No
3. Do you agree to compromise with fee at the cost of quality program/course?
 - a. Yes
 - b. No
4. Do you want extracurricular activities apart from studies?
 - a. Yes
 - b. No

5. Do you consider the accreditation of institute during admission like NBA, UGC, and AICTE?
 - a. Yes
 - b. No
6. Do you consider hostel facilities during admission?
 - a. Yes
 - b. No
7. Do you give preferences to health facility?
 - a. Yes
 - b. No
8. Do you consider transport facilities?
 - a. Yes
 - b. No.

9.2 Results

After doing the survey for checking the trust of students for a particular institute, we came to know that the student trust depends on many factors like fee, faculty, placement, hostel facilities, accreditation of the institute, and so on.

Based on this survey, we have drawn pie graph as shown in Fig. 1 that concluded the result and trusting percentage.

Figure 1 also shows that placement parameter has the highest priority of the students. And accreditation and faculty parameters are at the second and third preferences of students. Health and extracurricular activities are at fourth and fifth preferences of the students for selecting any particular institute. Further, we have computed ratings according to our defined mathematical formula (1) of a particular institute based on these ten (10) parameters. We have classified four institutes according to their excellent facilities provided by them. The first institute provides excellent placement and transport facility; the second institute provides best hostel facility; the third institute provides best extracurricular activities; and the forth institute provides best faculty members. Figure 2 shows ratings of institutes based on our defined parameters. From our survey observations, it has found that 90% students have rated to Institute 1 as 1 and computed trust value according to our defined formula (1) is 0.1 where 60% students have rated Institute 2 as 3 and its computed trust value is 0.3. It has also found that there are 70% students who rated to Institute 3 as 4 and calculated trust value is 0.4 where 80% students have rated to Institute 4 as 2 and computed trust value is 0.2. Thus, it is clear and validates that our formula (1) showed the accurate result as trust value is directly proportional to rating assigned by the students.

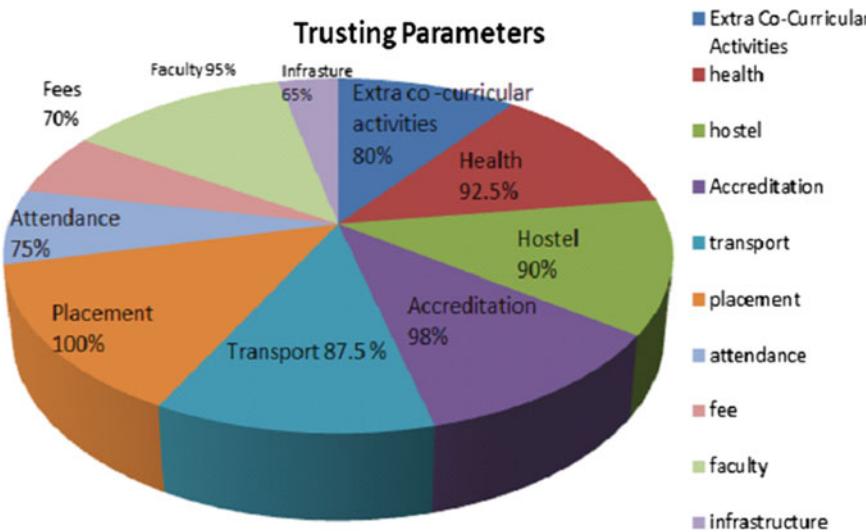


Fig. 1 Priorities of students with respect to various parameters

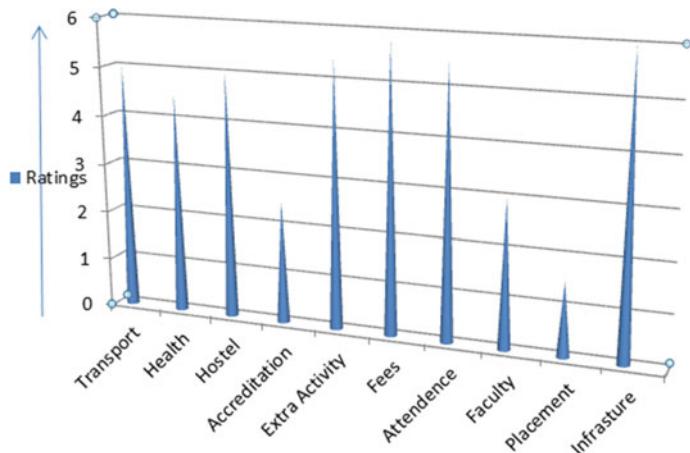


Fig. 2 Ratings of students with respect to trust parameters

10 Conclusion and Future Scope

Trust is very important in Web infrastructure. As the Web is increasing rapidly, the threat to authenticity of data is also increasing. Therefore, trust importance will also increase as time passes away. In this paper, we have proposed a mathematical novel trust scheme based on reputation approach for selecting an education institute for getting admission in particular course in Semantic Web. The accuracy of trust-

based approach ratings is more than simple ratings. Our proposed scheme has given significant results to compute trust as shown in Figs. 1 and 2.

In this paper, we have taken a survey of having 200 engineering students, which is not so big. Moreover, it is manual also. In future, we will develop automatic system to conduct survey of big data samples. We have planned to develop ontologies of education domain by using Semantic Web technologies to get more accurate results. We will design Web sites, and based on Web site rating, we will trust the Web site. Users will use and get admission in educational institute by checking the feedback given by the people who have already visited the Web site.

References

1. Al Jarhi, A., Gaaly, T.: Semantic, trust and proof in the semantic web. In: CSCI 585-Knowledge Engineering Spring, 2007
2. Golbeck, J., Parsia, B., Hendler, J.: Trust networks on the semantic web. In: Conference Paper in Lecture Note in Computer Science, January 2003
3. Girdhar, P., Kumar, S., Singh, M.: Comparative Study on Strengths and Weaknesses of Trust Algorithms In Semantic Web Based Networks. Springer, Berlin, Heidelberg (2013)
4. Wang, Y., Vassileva, J.: Trust and reputation model in peer-to-peer networks. In: Third International Conference on Peer-to-Peer Computing (P2P'03), p. 150, 2003
5. Mui, L.: Computational models of trust and reputation: agents, evolutionary games, and social networks. Ph.D. Dissertation, Massachusetts Institute of Technology (2003)
6. Massa, P., Avesani, P.: Trust-Aware Collaborative Filtering for Recommender Systems, vol. 3290. Springer, GmbH (2004)
7. Victor, P., De Cock, M., Cornelis, C., Pinheiro da Silva, P.: Towards a provenance-preserving trust model in agent networks. In: Proceedings of Models of Trust for the Web (MTW'06), Workshop at the 15th International World Wide Web Conference (WWW2006) (2006), ISBN 085432853X
8. Henze, N.: The Proof and Trust Layer of Semantic Web. IVS Semantic Web Group
9. Uszok, A., Bradshaw, J.M., Johnson, M., Jeffers, R.: Kaos policy management for semantic web services. IEEE Intell. Syst. **284**(5), 3241 (2004)
10. Kagal, L., Berners-Lee, T.: Where Policies Meet Rules in the Semantic Web. Technical Report, MIT (2005)
11. Caruana, S., Montebello, M.: Semantic Web Trust: The Next Step in Web Evolution
12. Golbeck, J.: Semantic Web Interaction Through Trust Network Recommender Systems. <http://trust.mindswap.org/ont/trust.owl>
13. Horrocks, I.: DAML + OIL: a reason-able web ontology language. In: Proceedings of EDBT, March 2002

Random-opposition-based Learning for Computational Intelligence



Divya Bairathi and Dinesh Gopalani

Abstract In this paper, random-opposition-based learning (ROBL) is proposed. ROBL is a generalized version of opposition-based learning (OBL). ROBL introduces randomness in OBL. ROBL is applied for some metaheuristics and artificial neural network. The examples are provided with preliminary results.

Keywords Computational intelligence · Opposition-based learning (OBL) · Random-opposition-based learning (ROBL) · Metaheuristics · Artificial neural networks (ANNs)

1 Introduction

Computational intelligence systems (CIS) are systems, capable of learning from experimental observations or data. Computational intelligence is used for those complex problems which are either hard or impossible to solve by traditional mathematical solving or stochastic in nature. CIS are generally based on nature-inspired computational methodologies. For example, artificial neural networks (ANNs) imitate the neural structure of the human brain [1]. An artificial neural network is a network of interlinked neurons and is capable of learning. Other examples are metaheuristics, reinforcement learning, opposition-based learning, etc. Metaheuristics are stochastic heuristics for optimization [2]. Most of the metaheuristics are inspired by nature and based on certain characteristics and behaviour of swarm, rules of physics, neurobiological systems, biological evolution and human behaviour. Some of the well-known metaheuristics are particle swarm optimization (PSO) [3], differential evolution (DE) [4], genetic algorithm (GA) [5], ant colony optimization (ACO) [6], evolution strategy (ES) [7], artificial Bee colony (ABC) [8], grey Wolf optimizer (GWO) [9], salp swarm algorithm (SSA) [10], etc. Reinforcement learning is based

D. Bairathi (✉) · D. Gopalani
MNIT Jaipur, Jaipur, India
e-mail: divyabairathjain@yahoo.co.in
URL: <http://www.mnit.ac.in>

D. Gopalani
e-mail: dgopalani.cse@mnit.ac.in

on human psychology. Opposition-based learning (OBL) [11] uses opposite numbers and imitates actions versus counter-actions, weights and opposite weights, and estimates and counter-estimates.

In this paper, random-opposition-based learning (ROBL) is proposed. Random-opposition-based learning is a generalized version of opposition-based learning. Basic ideas of opposition-based learning and random-opposition-based learning are discussed. ROBL is applied for some of existing learning algorithms. Examples are provided with experiments and results.

2 Opposition-based Learning (OBL)

Opposition-based learning (OBL) [11] founded on weights and opposite weights, estimates and counter-estimates, and actions versus counter-actions. The concept is based on opposite numbers. If upper and lower bounds of a range are defined by A and B , respectively, then, opposite of a real number Z is given by following equation.

$$\bar{Z} = A + B - Z \quad (1)$$

Opposition-based learning: Let Z be a real vector in range $[A, B]$, \bar{Z} be opposite of Z , $F(Z)$ be a function of concern and $G(Z)$ be a evaluation function (which calculates the objective value of $F(Z)$), then according to opposition-based learning, learning continues with Z if $G(F(Z))$ is better than $G(F(\bar{Z}))$, else with \bar{Z} .

3 Random-opposition-based Learning (ROBL)

In OBL, opposite number is calculated by computing counter-value in each direction. ROBL introduces randomness in OBL. Random opposite numbers (RONs) are computed instead of opposite numbers. In Fig. 1, positions of agents in two dimensions and three dimension are shown. The black point shows the original position of agent. All points' positions (Black, red, blue) are reachable by ROBL, whereas using OBL only red points are approachable. Random opposite numbers (RONs) are computed according to Eq. 2 for D -dimensional multidimensional problem.

```

for i = 1 to D
    if rand ≥ R
         $\hat{Z}_i = A_i + B_i - Z_i$           (2)
    end
end

```

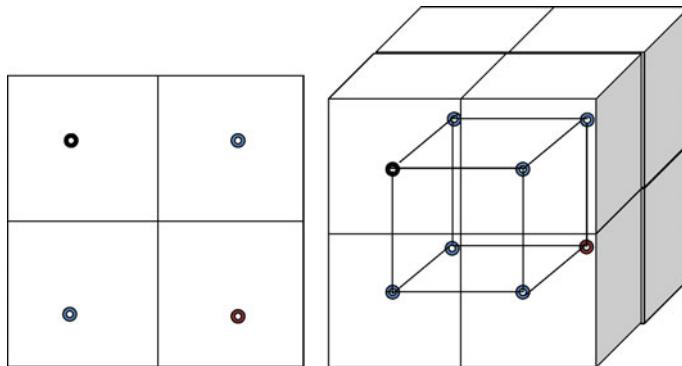


Fig. 1 Positions in two dimensions and three dimensions

Random-opposition-based learning: Let Z be real vector in range $[A, B]$, \hat{Z} be random opposite of Z , $F(Z)$ be a function of concern and $G(Z)$ be a evaluation function (which calculates the objective value of $F(Z)$), then according to random-opposition-based learning, learning continues with Z if $G(F(Z))$ is better than $G(F(\hat{Z}))$, else with \hat{Z} .

4 Random-opposition-based Metaheuristics

Metaheuristics are general purpose algorithms that provide acceptable solutions in a reasonable time for hard and complex optimization problems. High-level model of metaheuristics is as following.

```

Input: Initial solution  $s_0$ .
 $l = 0$  ;
 $s_l = s_0$  ;
Repeat
    Generate new solutions  $Q(s_l)$  from  $s_l$ ;
    Select solution from  $Q(s_l) \cup s_l$  to replace the current solution in  $(s_l)$ ;
     $l = l + 1$ ;
Until Stopping criteria satisfied
Output: Best approximate found so far.

```

For increasing exploration power, metaheuristics can be extended with random-opposition-based learning. Here, two algorithms SSA and PSO are selected for the extension. Both are swarm-based algorithms.

4.1 ROBL-based Extension of SSA

Salp swarm algorithm (SSA) [10] is inspired by the social behaviour of salps. In swarming behaviour, salps form salp chain. In this salp chain, the salps are connected with each other. The mathematical model of the salp chains concerns in two groups of salps: leader salp and follower salps. Leader salp is the foremost salp of the chain, and the remaining are considered as followers. Equations 3 and 4 model position update equations of leader and followers, respectively.

$$X_j^1 = \begin{cases} F_j + a_1 ((\text{UB}_j - \text{LB}_j) a_2 + \text{LB}_j) & a_3 \geq 0 \\ F_j - a_1 ((\text{UB}_j - \text{LB}_j) a_2 + \text{LB}_j) & a_3 < 0 \end{cases} \quad (3)$$

$$X_j^i = \left(X_j^i + X_j^{i-1} \right) / 2 \quad (4)$$

Here X_j^1 and X_j^i show the location of the first salp (leader) and i th salp (follower) in the j th dimension. F_j is the food source location in the j th dimension. UB_j and LB_j indicate the upper bound and lower bound of j th dimension, respectively. a_2 and a_3 are uniform randoms in range [0,1]. Value of a_1 is computed as follows.

$$a_1 = 2e^{-(4v/V)^2} \quad (5)$$

Here, v represents current iteration and V is max number of iterations.

Random-opposition-based extension of SSA (ROSSA): A new population (of same size as of original population) is created using random opposition in each iteration. After this, Elite solutions are selected from both populations to create a population of original size. Now, this population is treated as SSA population in further steps.

4.2 ROBL-based Extension of PSO

Particle swarm optimization is a metaheuristic-based social behaviour of fish schooling or bird flocking [3]. The agents are called particles. These particles move in search space using two equations: velocity update equation (Eq. 6)

$$v = v + c_1 r_1 (P_{\text{best}} - X) + c_2 r_2 (P_{\text{gbest}} - X) \quad (6)$$

and position update equation (Eq. 7).

$$X = X + v \quad (7)$$

Here, v and X represent particle's velocity and position, respectively. c_1 and c_2 are social and cognitive coefficients, respectively. p_{best} is global best particle position and P_{gbest} is best position of X found so far.

Random-opposition-based extension of PSO (ROPSO): For each particle, a random opposite particle is created using random opposition in each iteration. After this, Elite solution is selected from the two particles. Now, this solution is used in further iterations.

For analysis purpose, 25 benchmark functions are selected from [12, 13]. The benchmark numerical functions are given in Table 4. The algorithms ROPSO and ROSSA are tested for these benchmark functions and compared with PSO and SSA, respectively, in Table 5.

5 Random-opposition-based Artificial Neural Networks

An artificial neural network (ANN) is a computational intelligence system which imitates the neural structure of the human brain [1]. An ANN is a network of inter-linked neurons and is capable of learning. For many decades, artificial neural network (ANN) proves as a strong tool to efficiently deal with thousands of problems in many disciplines such as medical diagnosis, speech recognition, social network filtering, playing board and video games, machine translation, computer vision and in many other domains.

5.1 ROBL-based Extension of ANNs-ROANNs

A specimen of the ANN with random opposite weights is produced. A copy of entire network is created, and its weights are calculated as following.

```

for i = 1 to n
    if rand ≥ R
         $\hat{w}_i = a_i + b_i - w_i$                                 (8)
    end
end

```

Here, n is total number of weights in network. a and b are lower and upper bounds on weights, respectively. Error is calculated using Eq. 9 for both network. Network with lower error is chosen for next iteration.

Table 1 Datasets

Classification dataset	Attributes	Training samples	Test samples	Number of classes
Glass	9	163	163	2
Breast cancer	9	500	199	2
Heart	22	87	180	2

Table 2 ANN structure

Datasets	Number of links	Number of nodes	ANN structure
Glass	190	20	9-19-1
Breast cancer	190	20	9-19-1
Heart	1035	46	22-45-1

Table 3 Results

Dataset	Algorithm	Accuracy
Glass	ANNP	69.88 ± 1.19
	ROANN	70.02 ± 1.11
Heart	ANN	81.30 ± 0.41
	ROANN	81.97 ± 0.34
Cancer	ANN	96.55 ± 0.16
	ROANN	97.17 ± 0.17

$$\text{Error} = \sum_{n=1}^N |O_n - T_n| \quad (9)$$

Here, O_n = output of MLP, T_n = target output and N = number of test samples

For analysis purpose, three datasets from University of California, Irvine (UCI) [14] Machine Learning Repository: glass, breast cancer and heart are benchmarked. These benchmarks are used for classification. Specification for classification datasets is presented in Table 1. ANN structure used for experiment is shown in Table 2.

Results for datasets are shown in Table 3. Results show that algorithm ROANN performs significantly better than ANN.

6 Conclusion

In this paper, random-opposition-based learning (ROBL) is proposed. ROBL introduces randomness in opposition-based learning. Random opposite numbers are computed and used. Selection is based on elitism. ROBL is applied for metaheuristics

Table 4 Benchmark functions

Test problem	Function		Dim	Range	F_{\min}
Sphere	$F_1 = \sum_{i=1}^n x_i^2$		30	[-100, 100]	0
Schwefel 2.22	$F_2 = \sum_{i=1}^n x_i + [\prod_{i=1}^n x_{5rm}]^2$		30	[-10, 10]	0
Schwefel 1.2	$F_3 = \sum_{i=1}^n (\sum_{j=1}^i x_j)^2$		30	[-100, 100]	0
Max dimension	$F_4 = \text{Max}(x_{isteps} , 1 \leq i \leq n)$		30	[-100, 100]	0
Rosenbrok	$F_5 = \sum_{i=1}^n [100(2x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$		30	[-30, 30]	0
Shifted sphere	$F_6 = \sum_{i=1}^n (x_i + 0.5)^2$		30	[-100, 100]	0
De Jong 4	$F_7 = \sum_{i=1}^n i \cdot (x_i)^4$		30	[-5.12, 5.12]	0
Axis parallel hyper ellipsoid	$F_8 = \sum_{i=1}^n i \cdot x_i^2$		30	[-5.12, 5.12]	0
Exponential	$F_9 = -(\exp(-0.5 \sum_{i=1}^n x_i^2)) + 1$		30	[-1, 1]	0
Sum of different power	$F_{10} = \sum_{i=1}^n x_i ^{i+1}$		30	[-1, 1]	0
Step	$F_{11} = \sum_{i=1}^n (x_i + 0.5)^2$		30	[-100, 100]	0
quartic(noise)	$F_{12} = \sum_{i=1}^n i \cdot x_i^4 + \text{random}[0, 1)$		30	[-1.28, 1.28]	0
Schwefel	$F_{13} = \sum_{i=1}^n -x_i \sin(\sqrt{ x_i })$		30	[-500, 500]	-Dim* 418.9829
Rastrigin	$F_{14} = \sum_{i=1}^n [x_i^2 - 10\cos(2\pi x_i) + 10]$		30	[-5.12, 5.12]	0
Ackley	$F_{15} = -20\exp(-0.2\sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)) + 20 + e$		30	[-32, 32]	0
Griewank	$F_{16} = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(\frac{x_i}{\sqrt{i}}) + 1$		30	[-600, 600]	0
Alpine	$F_{17} = \sum_{i=1}^n x_i \sin(x_i) + 0.1x_i $		30	[-10, 10]	0
levy	$F_{18} = 0.1\sin^2(3\pi x_1) + \sum_{i=1}^n (x_i - 1)^2[1 + \sin^2(3\pi x_i + 1)] + (x_n - 1)^2[1 + \sin^2(3\pi x_n)] + \sum_{i=1}^n u(x_i, 5, 100, 4)$		30	[-50, 50]	0
Foxholes	$F_{19} = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^2 (x_i - a_{ij})^6} \right)^{-1}$		2	[-65, 65]	1

(continued)

Table 4 (continued)

Test problem	Function	Dim	Range	F_{\min}
Kowalić	$F_{20} = \sum_{i=1}^{11} \left[a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 - b_i x_3 + x_4} \right]^2$	4	[-5, 5]	0.00030
Six hump camel back	$F_{21} = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	2	[-5, 5]	-1.0316
Branin	$F_{22} = \left(x_2 - \frac{5}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8}\pi \right) \cos x_1 + 10$	2	[-5, 5]	0.398
goldstein price	$F_{23} = \left[1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 16x_1x_2 + 3x_2^2) \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)] \right]$	2	[-2, 2]	3
Hartmann-3	$F_{24} = - \sum_{i=1}^4 c_i \exp(- \sum_{j=1}^3 a_{ij} (x_j - p_{ij})^2)$	3	[1, 3]	-3.86
Hartmann-6	$F_{25} = - \sum_{i=1}^4 c_i \exp(- \sum_{j=1}^6 a_{ij} (x_j - p_{ij})^2)$	6	[0, 1]	-3.32

Table 5 Experimental Results

Function	ROSSA		SSA		ROPSO		PSO	
	Avg	STD	Avg	STD	Avg	STD	Avg	STD
F_1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
F_2	0.0000	0.0000	0.0000	0.0000	0.0176	0.0078	0.0378	0.0532
F_3	4.4540	3.6498	8.3847	4.2478	32.9389	14.4645	54.9748	34.9732
F_4	1.4556	0.7786	3.0731	2.0138	0.2430	0.0450	1.1237	0.2768
F_5	3.2634	4.8971	29.0042	0.1466	156.5643	45.8576	140.2671	84.0108
F_6	0.0342	0.0086	0.0676	0.0456	0.0003	0.0002	0.0003	0.0002
F_7	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F_8	0.0000	0.0000	0.0000	0.0000	0.0011	0.0006	0.0013	0.0008
F_9	-1	0.0000	-1	0.0000	-0.9998	0.0001	-0.9998	0.0001
F_{10}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F_{11}	0.0000	0.0000	0.0001	0.0001	0.1000	0.2000	0.2000	0.4000
F_{12}	0.0007	0.0004	0.0002	0.0020	0.2658	0.0786	0.1899	0.0698
F_{13}	-11858.872	634.3480	-12495.100	0.3567	-7578.757	7345.1453	-4872.520	7345.1453
F_{14}	0.5675	0.2478	0.6577	0.0876	36.8965	31.8374	49.8374	31.8374
F_{15}	0.0000	0.0000	0.0000	0.0000	0.2556	0.1543	0.4782	0.2348
F_{16}	0.0035	0.0016	0.0035	0.0032	0.0056	0.0076	0.0074	0.0059
F_{17}	0.0000	0.0000	0.0000	0.0000	0.0238	0.0343	0.0438	0.0453
F_{18}	0.0011	0.0005	0.0025	0.0005	0.0014	0.0009	0.0034	0.0045
F_{19}	1.0056	0.0055	1.6742	0.5743	2.4387	1.8493	3.6786	3.7867
F_{20}	0.0004	0.0002	0.0005	0.0003	0.0006	0.0004	0.0009	0.0006
F_{21}	-1.0316	0.0000	-1.0316	0.0000	-1.0316	0.0000	-1.0316	0.0000
F_{22}	0.3981	0.0007	0.3981	0.0006	0.3982	0.0001	0.3982	0.0002
F_{23}	3	0.0000	3	0.0000	3.6	0.7867	5.3	4.8973
F_{24}	-3.8598	0.0002	-3.8567	0.0019	-3.8545	0.0040	-3.8528	0.0047
F_{25}	-3.319	0.0000	-3.2505	0.1145	-3.2967	0.0585	-3.2454	0.0582

PSO and SSA. Artificial neural network is also extended with ROBL. The examples are provided with preliminary results. For future work, ROBL can be applied for other metaheuristics, different types of neural networks and reinforcement learning methods.

References

1. Hertz, J.: Introduction to the Theory of Neural Computation. Basic Books, vol. 1 (1991)
2. Glover, F.W., Kochenberger, G.A. (eds.): Handbook of Metaheuristics (Vol. 57). Springer Science, Business Media (2006)
3. Eberhart, R., Kennedy, J.: A new optimizer using particle swarm theory. In: Micro Machine and Human Science, 1995. MHS95., Proceedings of the Sixth International Symposium on. pp. 39–43, IEEE (1995)

4. Storn, R., Price, K.: Differential evolution a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* **11**(4), 341–359 (1997)
5. Holland, J.H.: Genetic algorithms. *Sci. Am.* **267**(1), 66–72 (1992)
6. Dorigo, M., Birattari, M., Stutzle, T.: Ant colony optimization. *IEEE Comput. Intell. Mag.* **1**(4), 28–39 (2006)
7. Rechenberg, I.: Evolution strategy: Natures way of optimization. In: Optimization: Methods and applications. possibilities and limitations, pp. 106–126. Springer, Berlin Heidelberg (1989)
8. Karaboga, D.: An idea based on honey bee swarm for numerical optimization, (Vol. 200). Technical report-tr06, Erciyes university, engineering faculty, computer engineering department (2005)
9. Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014)
10. Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M.: Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. *Adv. Eng. Softw.* (2017)
11. Tizhoosh, H.R.: Opposition-based learning: a new scheme for machine intelligence. In: Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on, vol. 1, pp. 695–701. IEEE (2005)
12. Ali, M.M., Khompatraporn, C., Zabinsky, Z.B.: A numerical evaluation of several stochastic algorithms on selected continuous global optimization test problems. *J. Glob. Optim.* **31**(4), 635–672 (2005)
13. Bansal, J.C., Sharma, H., Nagar, A., Arya, K.V.: Balanced artificial bee colony algorithm. *Int. J. Artif. Intell. Soft Comput.* **3**(3), 222–243 (2013)
14. Blake, C., Merz, C.J.: UCI Repository of Machine Learning Databases (1998)

Secure Key Agreement Model for Group Data Sharing and Achieving Data Deduplication in Cloud Computing



Manav Ashok Thakur, Santosh Bari, Rutuja Deshmukh and Smita Auty

Abstract Safe along with sound knowledge deduplication will noticeably cut back the communication and storage outlay in server storage space services, and has potential application in our huge data-driven society. Existing knowledge deduplication scheme area unit naturally intends to also resist brute-force attacks or confirm the strength and knowledge convenience. In the existing system, key written agreement is a drawback. In the proposed system, to get rid of key written agreement drawback, we tend to implement block style based mostly on key agreement protocol to share knowledge in CSP. It allows multiple partners to freely distribute information in cluster. The proposed system is a chunk base-type contract procedure to wire numerous partners, which might supply expand to amount of partners inside very CSP setting as per the construction of the chunk style, and to remove knowledge redundancy drawback we tend to use knowledge deduplication system within which knowledge owner can transfer file and send to cluster manager and cluster manager checks knowledge deduplication over native domain. In this, the approved person transfers knowledge over cloud environment to transfer file, and knowledge owner can send key request to key authority for secret key. Once the key is received from key authority owner can transfer file and send to cluster manager and check file deduplication on native domain and if file isn't offered on native domain then send file to cloud. During access, the user can send key request to any or all cluster members and after receiving the key from cluster member, file can be transferred.

Keywords Big data · Record distribution · Secure data reduplication · Unbiased imperfect chunk system

M. A. Thakur (✉) · S. Bari · R. Deshmukh · S. Auty
Pune, India

e-mail: manavthakur84@gmail.com

S. Bari
e-mail: sgbpune@gmail.com

R. Deshmukh
e-mail: rutuja.deshmukh13@gmail.com

S. Auty
e-mail: smitaauty@gmail.com

1 Introduction

Storage usage is probably going to extend in our massive information-driven society. While price of storage is comparatively low and advances in server storage solutions permit USA to store an increasing quantity of knowledge, there are units associated prices for the management, maintenance, process, and handling of such massive information [1, 2]. It is, therefore, unsurprising that efforts are created to reduce overheads by information duplication. The technique (of information of knowledge of information) reduplication is meant to spot and eliminate duplicate data, by storing solely one copy of redundant information. In different words, information deduplication technique will considerably reduce storage and information measure needs [3]. Users and information homeowners might not absolutely trust server storage suppliers; information (particularly sensitive data) area unit is probably encrypted before outsourcing. This complicates information deduplication efforts, as identical information is encrypted totally (different completely) users (or even constant user mistreatment different keys) can end in different ciphertexts [4, 5]. Thus, a way to expeditiously perform information deduplication on encrypted information could be a topic of current analysis interest. The system gives associate in nursing suitable space display place for citizens; however, this additionally publishes protection issues. In this condition, this is imperative on the way to confirm the protection to hold on information within the server. In [6–8], many systems were projected to conserve isolation of data information. On top of schemes solely thought of protection issues of one information holder. However, in a few systems many information homeowners really similar to firmly contribute to their data in a very cluster way. So, a procedure to chain safe cluster in order of distribution beneath CSP is required. A type disagreement procedure is in work to find an average consultation type for many partners to corroborate the safety of their later relations, and this procedure is often practical in CSP to hold safe and inexpensive in a row distribution. In cryptography, a key agreement protocol could be a protocol within which two or a lot of parties will agree on a key in such a simplest way that each influences the result. By using the key agreement protocol, the conferees will firmly throw and be given communication from solitary an extra abuse the frequent meeting input so as to they consent winning earlier. Purposely, a locked input concord code of deeds that the character cannot obtain the generated type by implementing malicious attacks, like eavesdrop. Consequently, the types of contract etiquette are frequently extensive in a job in interactive announcement environments by way of towering defense wants. All through this document, we contain a trend to gift associate in nursing economical and locked chunk-type contract by extending the constitution to hold many partners that permit many in turn homeowners to without stinting split the outsourced in order with elevated sanctuary and power. It is complete since the collect in turn division reproduction to maintain cluster in a row distribution in CS. Moreover, the etiquette resolves bid endorsement blunder acceptance goods.

2 Literature Survey

2.1 *Jean Liu, Benny Pinkas, Secure Encrypted Data Deduplication Without Additional Independent Servers, 2015*

Description: Secure encrypted data deduplication of knowledge while not further freelance servers encrypting knowledge on client-side before uploading it to server storage is crucial for safeguarding users' privacy. But client-side encoding is at odds with the quality following deduplication. Accommodative client-side encoding with cross-user deduplication is an energetic analysis topic. We tend to gift the rest secure cross-user reduplication theme that supports client-side encoding while not requiring any further freelance servers. Apparently, the theme relies on employing a password authenticated key exchange (PAKE) protocol. We tend to demonstrate that our theme provides higher security guarantees than previous efforts. We tend to show each the electiveness and also the efficiency of our theme, via simulation exploitation realistic datasets and an implementation.

2.2 *Maher Bellaire, Siam Keelveedhi, DupLESS: Server-Aided Encoded Redundancy Checking System, 2013*

Description: According to the author, the proposed architecture is providing security to duplicate data storage that resists against brute-force attacks and is known as DupLESS. In this, DupLESS client is using encrypting method for the message-based key which is actually received from server using PRF protocol. It will help client to store received encrypted data using existing service with platform of service deduplication, which helps to provide strong confidentiality. We can achieve good performance and can help to save space using deduplicated encryption storage nearer to service storage with the help of plaintext data.

2.3 *Shaik Mahabub Bashan, Enabling Storage Auditing in Server of Key Updates from Verifiable Outsource, 2016*

Description: In this document, the study on a way to source key updates for key exposure resilience using server storage auditing. It proposes the primary server storage auditing protocol for key updates using verifiable outsourcing. During this protocol, key updates square measure out sourced to the TPA and square measure clear for the consumer. Additionally, the TPA solely sees the encoded description

of the client's undisclosed input, because the consumer will additionally verify the encoded undisclosed type once downloading them from the TPA that supplies the formal security proof and also the performance simulation of the planned theme.

2.4 *V. Goutham, Enabling Server Using Key Exposure Resistance Storage Auditing, 2016*

Description: In the projected paradigm, it is deliberated on a way to contend with the client's key exposure in server storage auditing. A brand new customary is referred to as auditing protocol with key exposure resilience. The integrity of the information at one time keep in server will still be corroborated albeit the client's current secret key for server storage auditing is vacant in these styles of protocols. It is enacted in the definition and therefore the security model of auditing protocol with key exposure resilience, and has given the sensible answer. The protection proof and therefore the straight-line presentation assessment delineate that the protocol is secure and economical. The economical comparison between current protocol and earlier protocol supported BLS signature additionally has been provided.

2.5 *Emmanuel Cresson Olivier Chevassut, Provably Authenticated Group Daffier–Hellman Key Exchange, 2001*

Description: Group Daffier–Hellman protocols are used for authenticated key exchange (AKE). Which help to create pool of players each hold a shared secret key which is later on used, consider example, of multicast message integrity. As time goes, multiple different schemes are offered. But no company has taken an official action for cryptographic issue. In our paper, we suggest a security model to overcome this issue. By defining AKE as basic goal and another goal as entity-authentication. After that, we also suggest a model that helps to carry group authentication.

3 Existing System

In the existing system, uncountable conference key contract protocol is steered to secure electronic network conference. Most of them operate when all conferees are honest, however don't work once some conferees are malicious and commit to delay or destruct the conference. And the existing system does not support for deduplication. Earlier theme does not seem to be secure to share hint to cluster

and fail to attain information security and deduplication. However, projected system achieves each privacy conserving and free audit on cluster information.

The disadvantages of existing system is that, it is used when the popular of them work only what time all group member are honest and it do not work while a few group members are malicious and try to stop or destruct the discussion.

4 Projected System

In planned structure, building block design-based key agreement protocol which helps multiple participants and which may flexibly extend the amount of participants during a server setting in step with the structure of the block style and cut back information redundancy downside, we have a tendency to use information deduplication system. We have a tendency to develop a cross-domain primarily based system, within which we have to check multi-level deduplication for file uploading of our system, there is unit 2 domain users area unit out there. Once the user transfers a file, then the native manager can check file is existing already or not, and if file is already out there on the native domain, then file is not on hold and the native manager offers relation to the existing file. Once file is uploaded by file owner, file can be shared to all or any domain members. For sharing key to all or any members, we have a tendency to use block style primarily based on key agreement protocol. Exploiting this protocol, we have a tendency to divide a conference key into all or any participants and securely share information with cluster. For accessing any file to domain member, it should send key request to all or any member. After receiving key from all the members, each member will transfer the file. If any malicious user enters in the cluster and challenges to access bunch of information, then check is applied to send to TPA. The TPA checks the malicious user's details and removes the malicious user from cluster. An input conformity code of activities is in a job to get a characteristic discussion type for various members to verify the protection of their later transportation, and this etiquette is functional in CS to hold locked and cheap to run information giving out (Fig. 1).

The advantages of our system are as follows:

- The fiction building block answer deal practice that chains various applicants
- Gymnastically enlarge the quantity of accomplice in a CS according to the construction of the chunk design
- Data deduplication
- Key agreement protocol
- Privacy preserving
- Public Auditing.

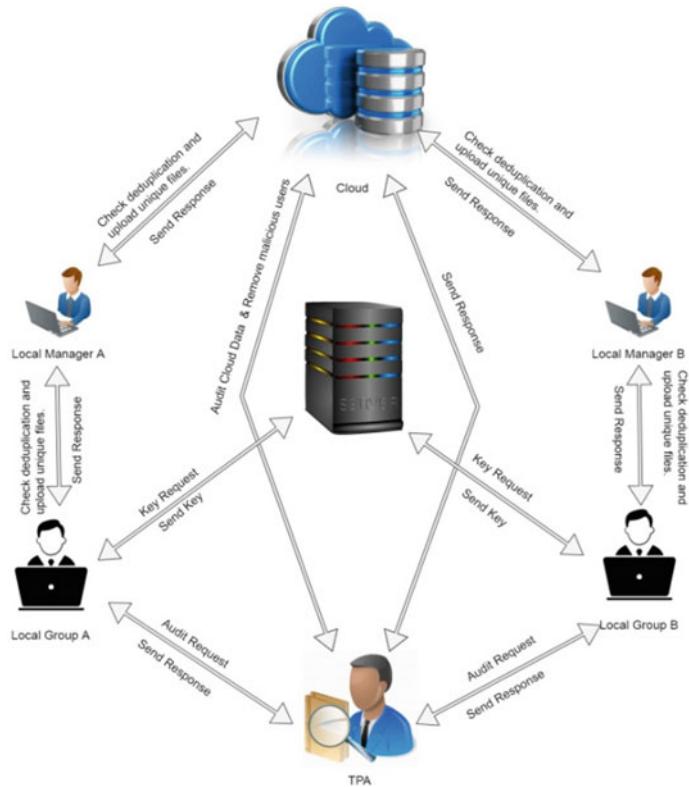


Fig. 1 A structure of design-based key agreement protocol

5 Algorithm

5.1 Algorithm 1: AES Algorithm Steps

- Step 1: Start the algorithm.
- Step 2: Then derive round keys set using the cipher key.
- Step 3: After that, initialize array using block data (plaintext).
- Step 4: Add initial round key at the array starting.
- Step 5: Perform the tenth and final state manipulation round.
- Step 6: Finally, copy encrypted data (ciphertext) as array is in final state.

6 Conclusion

The proposed system is a unique chunk system that supports cluster information distribution and deduplication theme to realize deduplication on server information. The multiple participants are often concerned within the protocol. During this project, domain manager and TPA play a necessary role in projected system. Domain or cluster manager will check deduplication at the time of file uploading, and TPA will audit on cluster sharing information and check if any malicious users square measure out there on cluster or not. If TPA notices any malicious activity in cluster, then the TPA can take away malicious users from the cluster. In the future work, we have a tendency to implement totally and shield the duplicate info from revealing, even by a malicious CSP, while not moving the ability to perform information deduplication.

References

1. Goutham, V.: Enabling server Storage Auditing with Key Exposure Resistance (2016)
2. Cresson, E., Chevassut, O.: Provably Authenticated Group Diffie-Hellman Key Exchange (2001)
3. Yu, J., Ren, K., Wang, C.: Enabling server storage auditing with verifiable outsourcing of key updates. *IEEE Trans. Inf. Forensics Secur.* **11**(6), 11 (2016)
4. Guo, H., Li, Z., Mu, Y., Zhang, X.: Cryptanalysis of simple three-party key exchange protocol. *Computers and Security* **27**(1–2), 1621 (2008)
5. Bresson, E., Chevassut O., Pointcheval, D.: Jean-Jacques Quisquater Provably Authenticated Group Diffie-Hellman Key Exchange
6. Liu, J., Pinkas, B.: Secure Deduplication of Encrypted Data Without Additional Independent Servers (2015)
7. Bellaire, M., Keelveedhi, S.: DupLESS: Server-Aided Encryption for Deduplicated Storage (2013)
8. Bashan, S.M.: Enabling Storage Auditing In server of Key Updates from Verifiable Outsource (2016)

Automated Road Surface Condition Monitoring System Using Machine Vision Technology



K. Sujatha, A. Ganesan, V. Karthikeyan, P. Sai Krishna, Shaik Shafiya, N. P. G. Bhavani and V. Srividhya

Abstract Monitoring of street surface conditions for different conditions like black-top, water, ice, and snow is required for most transportation offices are in charge of street support. Data on street surface conditions can be utilized to survey the requirement for upkeep benefit, think about the adequacy of various treatment techniques, and assess the nature of the support administrations conveyed by contractual workers crosswise over various upkeep yards. The road sensor utilizes three wavelengths and one photograph finder to decide the powers that are reflected from the street surface and is then ready to evaluate the street condition. By connecting this sensor to a GPS and a smaller than normal remote embedded Internet framework, the street conditions can be related to the right street position, making it conceivable to utilize the data in various applications.

Keywords Road condition · Information system · Server connection · Artificial neural networks · Map

K. Sujatha (✉) · V. Karthikeyan

Department of EEE, School of Electrical Engineering, Center for Electronics, Automation and Industrial Research, Dr. M.G.R. Educational and Research Institute, Chennai, India

e-mail: drksujatha23@gmail.com

A. Ganesan

RRASE College of Engineering, Chennai, India
e-mail: dragmephd@gmail.com

N. P. G. Bhavani · V. Srividhya

Meenakshi College of Engineering, Chennai, India
e-mail: sbreddy9999@gmail.com

P. Sai Krishna

Sree Rama Institute of Technology and Science, Kuppenakuntla, AP, India

S. Shafiya

Nalla Narasimha Reddy Education Societies's Group of Institutions, Hyderabad, AP, India

1 Introduction

Ongoing data on street surface conditions is additionally precious to the street clients who can utilize the data to enhance their movement and driving choices, for example, where, when, and in what mode to movement. At present, checking of winter street surface conditions is for the most part done through individual perceptions and manual chronicle, which is constrained in repeatability, subtle elements, and opportunities. While recent developments in sensor technologies had paved way for the use of continuous friction measurement equipment (CFME), Web-based surveillance video and spectroscopic snow and ice cover sensors have afforded new opportunities for quick and objective assessment of road surface conditions; they are costly for implementation and limited in spatial coverage and completeness. The proposed project is to further advance our new road condition monitoring solution featuring innovative applications of machine vision, artificial intelligence, and data fusion techniques on a platform of cloud-based wireless and Internet technologies.

2 Literature Survey

In the specialized report [1], there is a critical connection between car crashes and elusive street conditions. Grouping of tricky street conditions has been the subject of serious research for a few years. Online street conditions are monitored from the vehicle using Anti-Bolt Braking System (ABS), Traction Control System (TCS) with Electronic Stability Program (ESP) which helps to assess the dangerous situations and can be controlled using electronic gadgets. Be that as it may, dangerous street data would obviously intrigue all street clients. Thus, sending the tricky street data to a server would expand the data dispersal to something beyond one vehicle. For close infrared wavelengths of light, it has been demonstrated that the ranges of water, ice, and snow are recognizable [2]. This is the way in which four street conditions, dry, wet, frigid, and cold black-top, are used by a sensor for evaluation of dangerous street conditions. Today there are various optical model sensors for characterization of street conditions. All procedures in like manner abuse the distinction in force of the street pictures for various street conditions. Two strategies that do not utilize any extra brightening is a stereo camera framework joined with picture handling [3] and a strategy where the proportion of approaching and reflected light is estimated with two pyrometers [4]. The two systems are subject to road lights or approaching vehicles amid evening to work legitimately, which makes the strategies muddled. Notwithstanding, the expansive checking region of the stereo camera framework is a favorable position. A third procedure, the one that is utilized as a part of this examination, depends on laser diodes of various wavelengths and a photograph indicator [5–8].

3 Objective of the Work

- Automated machine-based review utilizing programming framework is more fast, precise, helpful, safe, and non-dangerous in correlation with conventional techniques.
- A gadget can be produced which is a mix of equipment and programming.
- Indian street territory conditions can be broke down and ordered in view of the parameters like normal power, length, width, length to width proportion.
- Analysis and classification of street territory are done outwardly and physically by human assessors. The choices taken by human auditors might be influenced by different outer variables which can be decreased by this mechanized framework.

4 Proposed Method for Road Condition Monitoring Using Image Processing

The initial development focus was toward testing the developed algorithms on secondary image data collected for other research projects. In the last two years, a purpose-built in-vehicle data collection system has been developed and is being improved upon. On the server end, an image processing application has been developed to communicate with in-vehicle units, store and process image data and show live status updated on the computed road condition status. The images of the road conditions are obtained from the video captured for different weather conditions. The force of the picture shifts concerning the idea of the street surface [9]. The highlights are separated and after that the street conditions are grouped utilizing feed-forward neural systems prepared with backpropagation algorithm (BPA). The diminished list of capabilities is utilized as a contribution to the ANN classifier lastly the order execution is approved with certain execution measures. Figure 1 demonstrates the schematic portrayal of the street condition observing framework.

5 Methodology

5.1 *Image Acquisition*

Video was captured for different road conditions. The continuous video is then segregated into frames using the video splitter software. The size of the images obtained for various road conditions is 281×228 . Any scientific activities can be performed just on square pictures. Consequently, the casings were resized to 200×200 . The stream is appeared in Fig. 2.

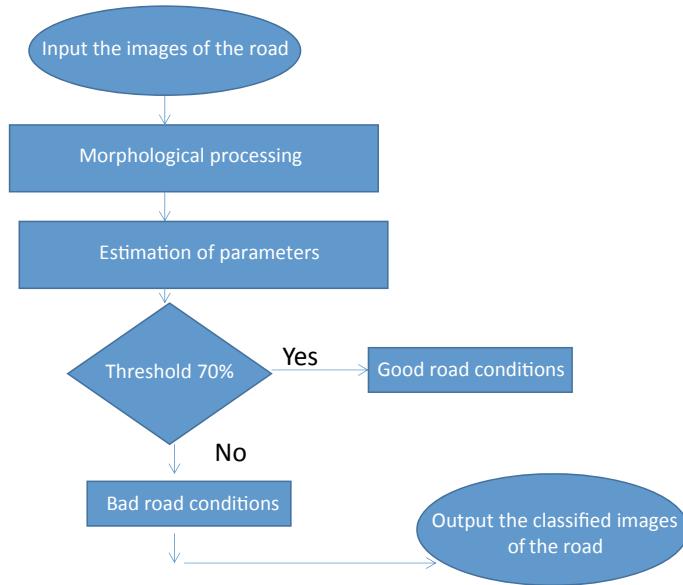


Fig. 1 Schematic for road condition monitoring

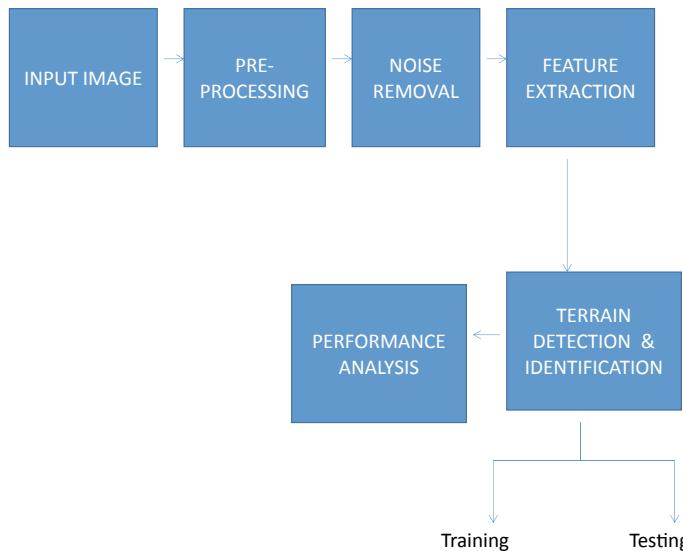


Fig. 2 Flowchart for road conditioning monitoring

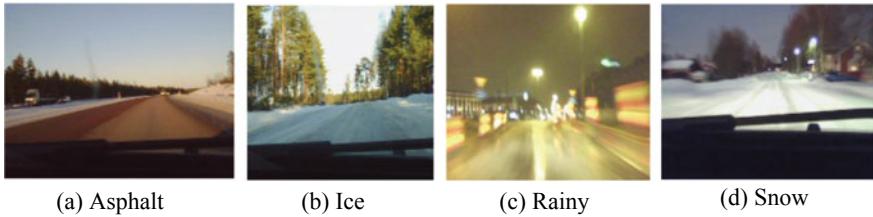


Fig. 3 Database for the road surface conditions for different weather conditions

5.2 Procedure for Data Collection

The street conditions relating to four distinctive climate conditions were gathered [2, 3]. The climate conditions relate to black-top, rain, ice, and snow separately. An aggregate of 120 pictures was assembled. For preparing and testing process, almost 80 and 32 pictures were considered. The example pictures relating to the above said four classes are appeared in Fig. 3.

5.3 Preprocessing

Preprocessing is done to expel commotion. The casing size obtained was 320×240 . It is adequate that we take 30×30 bit of the picture for encourage investigation. The extricated fire pictures if debased with clamor, at that point separating should be finished. The different channels incorporate unsharp cover channel, maximum channel, minimum channel, variance channel, mean channel, and median channel. The sifted pictures utilizing the above said channels are shown in Fig. 4 demonstrated as follows. It is derived that the execution of middle channel on clamor expulsion was great when contrasted and alternate channels. The mean channel likewise performs separating to a specific degree yet the clearness of the sifted picture is somewhat low when contrasted and the middle channel. If the channel is corrupted with noise, then the quality of picture is lost. Consequently, it is smarter to utilize a middle channel for commotion evacuation relating to this circumstance. The reproduction for expansion and evacuation of clamor was finished utilizing Image J. The execution assessment for clamor evacuation is finished utilizing mean squared error (MSE) and peak signal-to-noise ratio (PSNR) in decibel (dB) and the outcomes are arranged in Table 1. The MSE is least and PSNR esteem is most extreme if there should arise an occurrence of the middle sifting. Thus, it is induced that middle channel is viable if there should be an occurrence of expelling the salt and pepper commotion. The $M \times N$ is the span of the picture, $O(m, n)$ indicates the first picture and $R(m, n)$ signifies the recovered picture. The articulations for MSE and PSNR are given by Eqs. 1 and 2 individually.

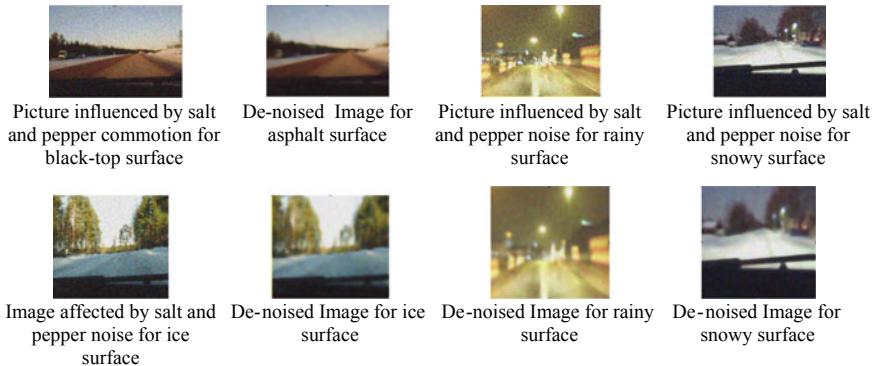


Fig. 4 Outputs for filtering the noise

Table 1 Performance of filtering action

Filter	MSE	PSNR (dB)	S.No	Filter	MSE	PSNR (dB)
Minimum	1.145	47.54	4.	Maximum	0.83	48.94
Variance	2.228	44.65	5.	Mean	0.146	56.49
Unsharp mask	4.327	41.77	6.	Median	0.0038	62.33

$$\text{PSNR} = 10 * \log_{10}[(255)^2 / \text{MSE}] \quad (1)$$

$$\text{MSE} = \sum_{m,n} \sqrt{[O(m, n) - R(m, n)] / (M * N)} \quad (2)$$

5.4 Features Extracted

5.4.1 Highlights

The highlights are the essential example show in the picture that gets reshaped in different ways. Highlights like region of the fire, mean force, most extreme power, least force, kurtosis, centroid, mode, border, standard deviation, and middle for every one of the gatherings of fire pictures are extricated utilizing Image J (Table 2). The definitions for the different highlights are as per the following.

5.4.2 Area

Area is the pixel area to the interior of the object defined as the sum of total number of pixels inside the boundary and pixels on the boundary.

Table 2 Feature extraction

S. no	Road condition	Image of the road	Area	Mean	Standard deviation	Mode	Min	Median
1.	Asphalt		28,634	54.239	36.608	19	16	46
2.	Ice		23,240	165.437	30.395	180	88	166
3.	Rain		27,900	80.276	53.734	21	17	86
4.	Snow		36,287	97.523	87.639	23	17	38

5.4.3 Mean

The mean of an informational index is essentially the math normal of the qualities in the set, got by summing up the qualities and isolating the quantity of qualities.

5.4.4 Mode

The mode of a set of intensity values is the value in the set that occurs most often.

5.4.5 Standard Deviation

It is interpreted as how the numbers (intensity values) are spread out.

5.4.6 Maximum and Minimum Intensity

It is defined as the highest and lowest values of the intensity in the flame image.

5.5 Training and Testing of BPA

The arrangement is finished utilizing feed-forward neural system utilizing BPA for preparing process. The different kinds of BPA utilized are Quasi-Newton and Resilient BPA.

6 Results and Discussion

Figure 5 denotes the histogram analysis. Tables 3 and 4 denote the features extracted and the classification efficiency of ANN. Figure 6 and Table 5 show the results for ANN training using MATLAB.

7 Conclusion

It is clear that the nature of ignition can be observed utilizing wise strategies like fake neural systems by examining the street pictures caught. The different kinds of BPA are utilized for preparing, testing, and approval. The execution measures are observed to be inside acceptable cutoff points. Street business people could utilize this data while disseminating salt. The advantage would be that the trucks would just disperse salt where it is required, i.e., where there is ice. This street condition

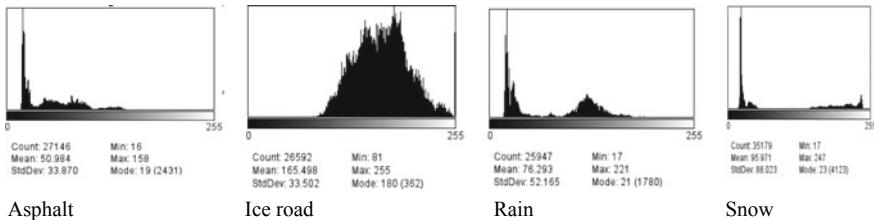


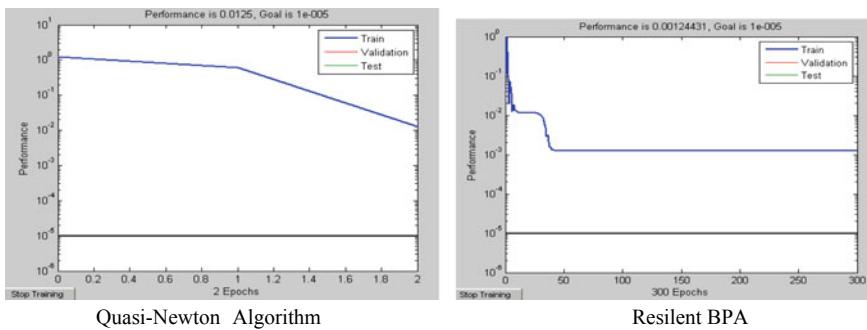
Fig. 5 Histogram analysis

Table 3 Training the ANN using the extracted features

Name of the algorithm	No. of iterations	Three input features	Four input features	Five input features	Six input features
		MSE	MSE	MSE	MSE
Quasi-Newton algorithm	211	2.54×10^{-5}	3.66×10^{-5}	1.6×10^{-5}	2.89×10^{-5}
Resilient BPA	201	2.41×10^{-5}	3.18×10^{-5}	1.17×10^{-5}	2.37×10^{-5}

Table 4 Classification efficiency for various algorithms

Class	No. of samples for training	No. of samples for testing	No. of samples for validation	Classification efficiency (%)	
				Quasi-Newton	Resilient BPA
1	20	8	2	72.98	98.33
2	20	8	2	63.21	99.13
3	20	8	2	63.21	98.67
4	20	8	2	63.21	98.23
Total/average	80	32	8	65.6	98.71

**Fig. 6** Results for ANN-based training using the extracted features**Table 5** Target versus actual values for training the ANN

Road condition	Target values	Actual values obtained for various types of BPA	
		Quasi-Newton	Resilient BPA
Asphalt	0.1	0.2500	0.1499
Rainy	0.2	0.2500	0.1501
Icy	0.3	0.2500	0.3003
Snow	0.4	0.2500	0.4001

data additionally shows when the salt water solidifies, so business visionaries get a sign when it's the ideal opportunity for redistribution. Road clients could design their excursions in light of how the street conditions are right now. The data could be utilized as driver alarm. The vehicles are constructed with information gathering and transfer systems. Both the driver and also dynamic security frameworks in the vehicle could utilize this data. Road climate figures would likewise profit of comprehending what medium is available out and about for their elusive street conditions estimations.

References

1. Sujatha, K., Pappa, N.: Combustion monitoring of a water tube boiler using a discriminant radial basis network. *ISA Trans.* **50**, 101–110 (2011)
2. Sujatha, K., Pappa, N.: Combustion quality estimation in power station boilers using median threshold clustering algorithms. *Int. J. Eng. Sci. Technol.* **2**(7), 2623–2631 (2010)
3. Bhavani, N.P.G., Sujatha, K.: Soft sensor for temperature measurement in gas turbine power plant. *Int. J. Appl. Eng. Res.* **9**, 21305–21316 (2014)
4. Sujatha, K., Pappa, N., Senthil Kumar, K., Siddharth Nambi, U.: Monitoring power station boilers using ANN and image processing. In: *Advanced Materials Research*, vol. 631–632, pp. 1154–1159. Trans Tech Publications, Switzerland (2013)
5. Sujatha, K., Pappa, N., Senthil Kumar, K., Siddharth Nambi, U., Raja Dinakaran, C.R.: Intelligent parallel networks for combustion quality monitoring in power station boilers. In: *Advanced Materials Research*, vol. 699, pp. 893–899. Trans Tech Publications, Switzerland (2013)
6. Sujatha, K.: Flame monitoring in power station boilers using image processing. *ICTACT J. Image Video Process.* **02**, 427–434 (2012)
7. Sujatha, K.: Combustion quality estimation in power station boilers using SVM based feature reduction with Bayesian. *Eur. J. Sci. Res.* **120**(2), 189–198 (2014). Elsevier
8. Sujatha, K., Senthil Kumar, K., Godhavar, T., Ponmagal, R.S., Bhavani, N.P.G.: An effective distributed service model for image based combustion quality monitoring and estimation in power station boilers. In: *Advanced Computer and Communication Engineering Technology, Lecture Notes in Electrical Engineering (LNEE)*, vol. 362, pp. 33–49 (2016)
9. Sujatha, K., Pappa, N., Senthil Kumar, K., Siddharth Nambi, U., Raja Dinakaran, C.R.: Automation of combustion monitoring in boilers using discriminant radial basis network. *Int. J. Artif. Intell. Soft Comput.* **3**(3), 257–275 (2013)

Evolution of Malware and Its Detection Techniques



Sanjay K. Sahay, Ashu Sharma and Hemant Rathore

Abstract In today's world, information is one of the most valuable assets, but there is a major threat to it by the evolving second-generation sophisticated malware, because it can enter the networks, quietly take the confidential data/information from the computational devices, and can cripple the infrastructures, etc. To detect these malware, time-to-time various techniques are proposed. These methods range from the early day signature-based detection to machine/deep learning techniques. Therefore, to understand the evolution of malware and its detection technique, this paper presents an overview of the evolution of malware and it's detection techniques. It discusses in details the various type of second-generation malware and the popular detection techniques used to detect it, viz. signature matching, heuristic methods, normalization, and machine/deep learning techniques.

Keywords Malware · Polymorphic · Metamorphic · Machine learning

1 Introduction

Despite the advancement in the development of anti-malware, the number of attacks from malware and is on an uptrend. An estimate by Symantec shows that the rate of creation of new instances of malware was 41%, with a total of over 400 million existing new malware instances [1]. According to FireEye survey [2], 47% of the organization experienced malware security incident/network breaches. Internet security threat report of 2017 says that there is a 36% increase in the attack with each day ~4500 new attacks to the victim's data [3]. As per McAfee technical report of

S. K. Sahay (✉) · A. Sharma · H. Rathore
Department of CS & IS, BITS, Pilani, Goa Campus, Goa, India
e-mail: ssahay@goa.bits-pilani.ac.in

A. Sharma
e-mail: p2012011@goa.bits-pilani.ac.in

H. Rathore
e-mail: hemantr@goa.bits-pilani.ac.in

2014, there were more than 200 million known malware samples [4] and the year 2013 was the mega breach year [5]. In the last quarter of 2015, a 26% increase in new ransomware samples has been reported [6], and in the first quarter of 2016, the Quick Heal Threat Research Lab received more than 340 million malware samples running into hundreds of thousands of devices [7]. Symantec reported 54 zero-day vulnerabilities, and it is doubling each year [8].

To counter/defend the malware, there are many anti-malware defense systems based on signature matching, heuristic analysis, normalization, and machine/deep learning. Some malware can be detected easily by signature matching anti-malware software. But the use of advanced obfuscation techniques, signature matching technique can't detect advanced/second generation or unknown variant of malware. However time to time, a number of static and dynamic methods have been proposed [9–15]. But, it appears that so far proposed approaches are not sufficient to detect the complex second-generation malware to limit or prevent the damages [16] in the fast-growing Internet and computational devices usage into our daily life. Hence, computational device security is viewed as one of the most important areas to be addressed.

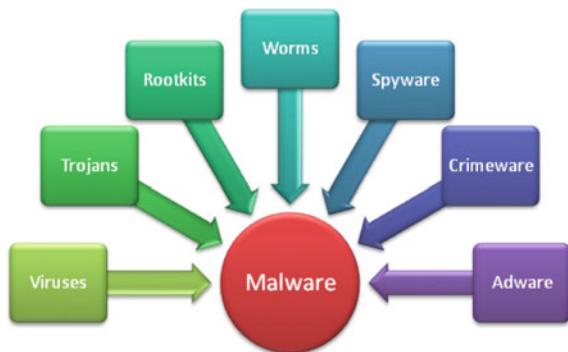
The first malware was created for fun, but now its a profit-driven industry [17] and in the last couple of years advanced malware uses complex obfuscation methods, viz. control/data flow permutation, compression, heap spray, etc. [18] to evade the detection techniques. To detect these malware, various methods range from the early day signature-based detection to machine learning techniques [16] are proposed. But it appears that malware developer is ahead of anti-malware group by using sophisticated obfuscation techniques and applying new tactics to evade the anti-malware. Therefore, to know the development of advanced malware and its detection technique in this paper, we discuss the evolution of malware and it's detection techniques. It discusses in details the type of second-generation malware and the popular detection techniques used to detect it, viz. signature matching, heuristic methods, normalization, and machine/deep learning techniques.

Rest of the paper is structured as follows: Sects. 2 and 3 present a survey on the different types of malware and its detection techniques respectively. In Sect. 4, we summarize our conclusion.

2 Types of Malware

The first generation, i.e., static malware are generally classified on the basis of their infection strategy as viruses (*it attaches itself to an executable and propagates the copies to other programs*), worms (*an executable that propagates copies to other computers*), and trojans (*a program that contains unexpected additional functionality*) [19]. Few other notable first-generation malware are root-kits, spyware, crime-ware, adware, etc. (Fig. 1) [20]. They all exhibit different sort of malicious behavior on the target systems, but design of the malware remain unchanged. But in the second generation, i.e., dynamic malware after each infection, structure of the malware changes to create a new variant keeping the intent same [16]. On the basis of the mechanism

Fig. 1 First-generation malware



by which either the code or the structure are obfuscated to conceal the signature of the malware, the second-generation malware can be further classified as encrypted, oligomorphic, polymorphic, and metamorphic malware.

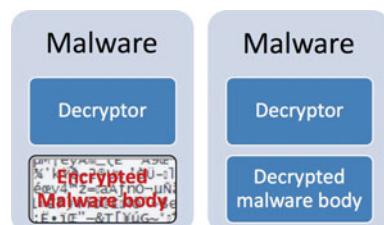
Encrypted Malware

Encryption is the first concealment technique in which malware body consists of a encrypted malicious code, key, and encryption/ decryption algorithm (Fig. 2). In this, body of the malware is XORed with the generated key to make it difficult to detect. The main objective to create encrypted malware was to evade the static code analysis and traditional signature-based detection technique. It infects the system by decrypting itself using decryption algorithm and a key, after that it again encrypt itself by the encryption algorithm and generates a new key for another variant to avoid the detection mechanism. However, encrypted malware can be detected by analyzing the encryption/decryption algorithm because the decryption/encryption algorithm does not change, e.g., Cascade was the first encrypted malware [21], later on applying this method Zombie/Win95 and Wind95/Mad were designed.

Oligomorphic Malware

Limitation of the encrypted malware, i.e., invariance of the decryptor in the particular malware variants made easy for the anti-malware to detect it by simply finding the signature of the decryptors. Hence, led to the development of various concealment

Fig. 2 Encrypted malware



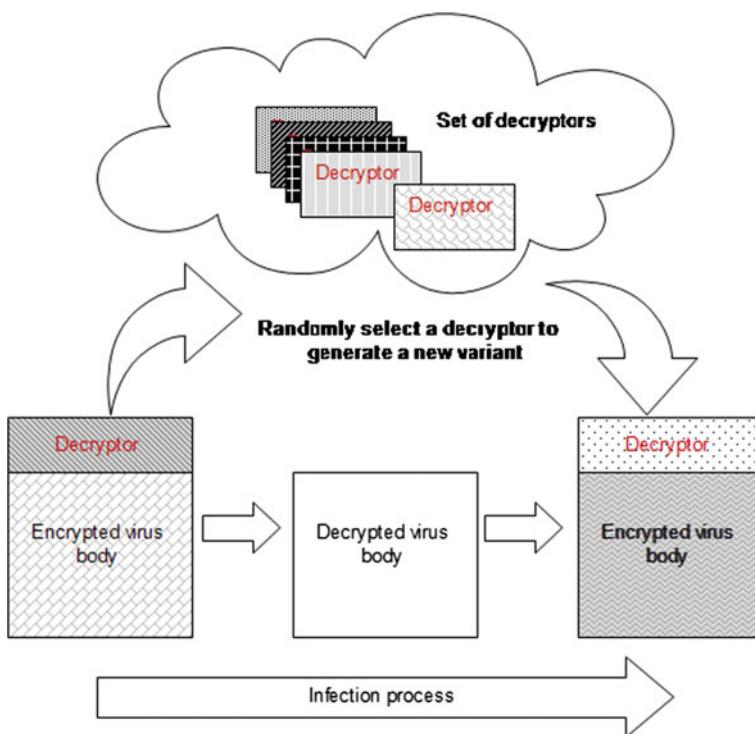


Fig. 3 Oligomophic malware

techniques to evade the detection mechanism. In this malware (Fig. 3), decryptors are mutated in the malware variant; i.e., it provides set of obfuscated decryptors. However, initially this malware was able to change the decryptor slightly [22], and at most can create few hundreds of decryptors [23], e.g. Win95/Memorial had the ability to build 96 different decryptor patterns [23]. To detect oligomorphic malware, traditional signature matching methods can be applied by obtaining the signature of the decryptors. However, generally, signature-based techniques are not recommended for the detection of oligomorphic malware [24].

Polymorphic Malware

Polymorphic malware is similar to oligomorphic malware, but it can generate millions of decryptors by mutating the instructions in the variant of the malware [25] to evade the signature matching detection technique. In this malware, mutation engine generates an encryption algorithm and a corresponding decryption algorithm, then malware code and mutation engine both get encrypted to generate a new variant of a particular malware. Figure 4 shows the structure of a polymorphic malware which has two parts, a decryptor and the body of the malware. To infect the system, mutation engine generates a new decryptor and get attached to the body of the

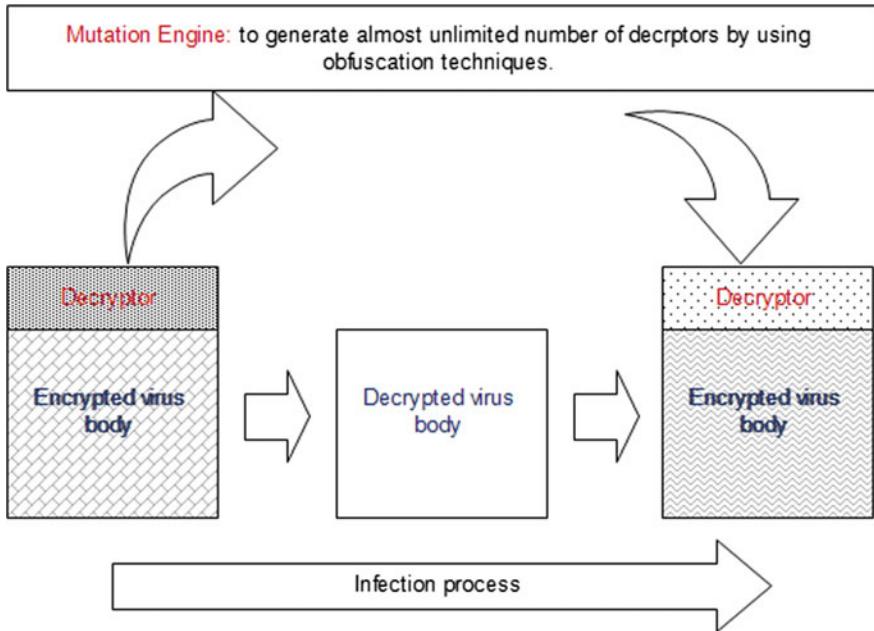


Fig. 4 Polymorphic malware

malware to create another variant of the malware [16]. These malware are created by using different obfuscation techniques, viz. register renaming, subroutine permutation, transposition, changing the control flow, subroutine inlining and outlining, equivalent code substitution, and dead code insertion [22], e.g., 1260 was the first polymorphic malware created by Mark Washburn in 1990 [24]. Although it can generate a large number of different decryptors, traditional signature-based detection mechanism can be applied to detect the polymorphic malware by simply finding the original program using the emulation techniques [22].

Metamorphic Malware

In Metamorphic malware (Fig. 5), instead of mutating the decryptors the malware body is mutated itself (i.e., body-polymorphic) to create a new variant without changing its actions to evade the detection [24]. To create the variants similar to polymorphic malware, various obfuscation methods, viz., dead-code insertion, data modification, control/data flow modification, register renaming, subroutine permutation, and equivalent code substitution are used to get the metamorphic behavior. These malware are nearly impossible to detect by the traditional signature-based techniques and hence in future will bring damages in large scale to computational devices. However, without increasing the size creating a metamorphic malware is a challenging task. In 1998, the first metamorphic malware Win95/Regswap was created, thereafter in 2000 Ghost/Win32 was made with 3,628,800 variants [26], and in 2001 using NGVCK

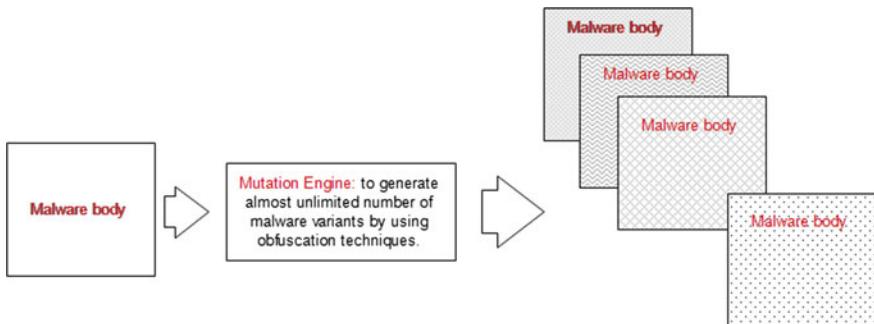


Fig. 5 Metamorphic malware

one of the strongest W32/NGVCK metamorphic malware was designed. Some of the malware which exhibits true metamorphic behavior are Second Generation Virus Generator [27], Phalcon/Skism Mass-Produced Code Generator [21], Virus Creation Lab and Mass Code Generator for Win32 [28].

3 Detection Techniques

Time-to-time various techniques have been proposed/used to detect the malware. These methods range from the early day signature-based detection to machine/deep learning techniques [9–13, 16]. Initially, the anti-malware was basically based on the assumption that the malware structure remains same, but the advanced second-generation malware structure changes after each infection. Currently, anti-malware (antivirus) companies like Symantec, McAfee, Kaspersky, and Fire eye provide major defense against malware attacks to the normal users. However, the exponential increase in malware numbers and use of advanced obfuscation techniques help the malware to evade the traditional signature-based anti-malware software. Also, as discussed in Sect. 2 that the variant of second-generation malware very much differs to each other. Therefore, its a need of time that academia as well anti-malware companies should continuously and jointly work together to prevent damages from the evolving advanced complex malware. Thus, we discuss below the different popular techniques that are used to detect the malware.

Signature-Based Detection

The traditional signature-based detection (Fig. 6) is an effective and simple technique to detect the known malware [29]. In this technique after identifying the malware, a unique short sequence/pattern of bytes are extracted to differentiate the malware from the benign programs [30]. These signatures are selected long enough to represent a specific malware, e.g. *klez.E/Worm* and *Worm/MyP-arty.A* signatures are 33be732d4000bd08104000e89eeffff80bd08104000be7d2d4000e849eaffff6a00e8350

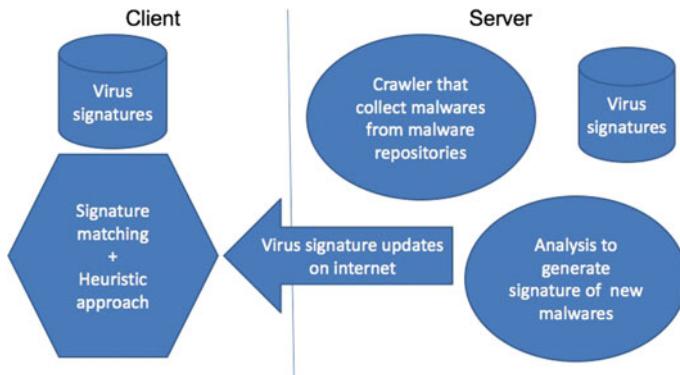


Fig. 6 Traditional detection system

0000064756d6d792e65786500653a5c77696e646f77735c53795374656d33325c644c6c
63616368655c6464642e65786500ff254c404000ff25544040a, and aa328cf24554d90b3
07c407eca9a4cf02a4d5a90000332c8b26904ffffb840f97f370080040e1fbae00b409cd2
1b8014c001f027c54686973c363616e042568d54562e2c876b0ffbf0420444f53 respectively [31]. In this method, mostly the files in the system are scanned to find a defined signature of the malware. But any slight mismatch in the malware will escape the detection. Therefore, the use of wildcard was proposed by Veldamna and Wu-Manber to detect the slight variance in the malware [25]. Hence, there is a finite probability that the wildcard method may detect the advanced malware, e.g., W32/Regswap [25].

Signature-based detection is a completely reactive technique. Therefore, these techniques are not able to defend the attacks/threats from the new or previously unseen malware until it causes the damages. Hence, signature-based techniques must be replaced by other robust methods to detect the second-generation malware [32]. In this, generally after the attack, domain expect (antivirus companies) get a copy of malware, and then they analyze to create a new signature of it. Thereafter, the signature database is updated to their antivirus clients. The whole process is a slow and time-consuming process. According to Hu et al. [33], on an average, it takes 54 days to release the signature of the malware. A study by Damballa [34] shows that around 15% of samples are still undetected even after 180 days.

Heuristics-Based Detection

Between 2000–2010 heuristic-based detection technique combined with signature-based detection was a major defense against malware, and heuristic method was one of the promising approaches to detect the new or previously unseen malware [35]. In this method, two approaches are used for the identification. Firstly, in static methods suspicious programs are analyzed to obtain a defined pattern in the program, if any, and if the result crosses the threshold, then the file is said to be infected [36]. Secondly, in dynamic methods, emulation approaches are used, viz. simulating the processor and OS to detect suspicious operations on a virtual machine. This dynamic

approach is generally prone to false alarm [28]; hence, system will be vulnerable by considering the false alarm as a real malware. Therefore to minimize the false alarm, the results of this detection technique are augmented with other methods [37]. Also, in this technique, domain experts based on malware analysis determine rules which can separate malware from benign. Rules should be flexible enough to identify malware and its variants with high accuracy and low false alarm rate. But again, these rules are manually developed by domain experts which is a slow and time-consuming process.

Malware Normalization

By using the sophisticated obfuscation techniques, malware developers have developed automated advanced malware generation toolkits like Zeus, Ultimate Packer for Executable, and Mitsfall [38]. These kits can generate thousands of malware in a day which is extremely hard to detect with signature or heuristic-based malware detection techniques. In this method, normalize executable/ malware are obtained by removing the obfuscation in the program and then it can be used to increase the detection accuracy of an existing anti-malware (Fig. 7) [39].

Christodorescu et al. [40] proposed the normalization of the obfuscation techniques, viz. code reordering, packing, and junk insertion [41] and improved the accuracy of many anti-malware. In this, later on, a generalize normalization of the malware has been proposed by Armor et. al. [42] that can keep the obfuscations in the form of automata structures and use them to normalize the advanced malware. Also, Armoun et al., in 2014, proposed a normalizer which can detect the variant of malware up to 81% accurately by storing many obfuscation techniques in the form

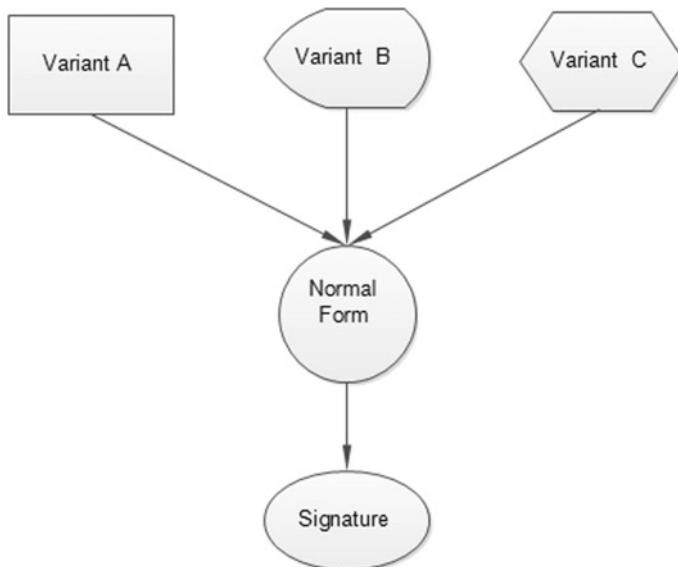


Fig. 7 Malware normalization

of automata structures to normalize the metamorphic malware [42]. One important advantage of malware normalization is that if normalization is done properly, it can detect an unknown variant of a particular malware or previously unseen malware.

Machine/Deep Learning Techniques

Nowadays, machine learning [43] techniques are widely explored to detect the unknown or previously unseen malware. This technique not only detects the known malware but can also detect the unknown malware by getting knowledge from the previously detected malware. The technique is a two-step process, in the first step feature (e.g., API Calls, N-gram, Strings, Opcodes, Control Flow Graph, etc.) are extracted from the know datasets which plays a vital role, not only to represent the target concept but also to speed-up the learning and classification/detection processes. In the second step, appropriate machine learning techniques, viz. Decision Tree [44], Naive Bayes [45], Data Mining [46], Hidden Markov Modes [47], and Neural Networks [44] are trained for detection/classification of malware.

Recently, deep learning has been showing promising results in various artificial intelligence applications like image recognition, natural language processing, language modeling, and neural machine translation. Although, in general, it is computationally more expensive as compared to classical machine learning techniques, their results are found to be more effective in some cases. In this, artificial neural network (ANN) is a mathematical model which simulates a subset of highly complex, nonlinear, parallel structure of human brain [48]. Deep neural network is an ANN with multiple hidden layers simulating the human brain. In DNN, first layer works as input layer and final layer works as the output layer. Input layers output works as input for the first hidden layer which performs some mathematical operation, then the nonlinear transformation output is fed to next layer and so on. When ANN is used as a classifier, final layer produces classification result. Large ANN (which contain multiple hidden layers) often suffers from local minima problem, and thus, it may take a longer time to train them through back-propagation. Convolutional neural network (CNN) is a special DNN with a two-layered architecture (convolutional layer and polling layer) used for image recognition or pattern matching. Recurrent neural network (RNN) is again a special DNN with a different memory structure. RNN has given good results in the field of language processing or speech recognition [49].

4 Conclusion

In today's information era, most of the computational devices are connected to the Internet, as a consequence, these devices are very much vulnerable to the cyber threat/attack from the advanced malware. It can enter the networks, quietly take the confidential data/information from the computational devices, and can cause a problem to the infrastructures etc. To defend the attacks/threat from the malware, time-to-time anti-malware has been developed. But it appears that malware developer is ahead of anti-malware group by using sophisticated obfuscation techniques

and applying new tactics to evade the anti-malware. Therefore, to know the development of advanced malware and its detection technique in this paper, we discuss the evolution of malware, and it's detection techniques. We discussed in details the type of second-generation malware and the popular detection techniques used to detect it, viz. signature matching, heuristic methods, normalization, and machine/deep Learning techniques.

References

1. Internet Security Threat Report (ISTR): Technical report. Symantec Corporation, April 2012. Date last accessed 31 May 2018. http://www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_2011_21239364.en-us.pdf
2. The Need for Speed: 2013 Incident Response Survey. Technical report, FireEye (2013). Date last accessed 31 May 2018
3. Internet Security Threat Report (ISTR). Technical report, Symantec Corporation, April 2017. Date last accessed 31 May 2018. <http://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf>
4. McAfee Labs Threats Report: Technical report, McAfee, June 2014. Date last accessed 31 May 2018. <https://www.mcafee.com/us/resources/reports/rp-quarterly-threat-q2-2014.pdf>
5. Internet Security Threat Report (ISTR): Technical report, Symantec Corporation, April 2014. Date last accessed 31 May 2018. <https://www.symantec.com/content/dam/symantec/docs/reports/istr-22-2017-en.pdf>
6. Daly , M.K.: Advanced persistent threat. USENIX 4 (2009)
7. Quick Heal Quarterly Threat Report Q2 2015: Technical report, Quick Heal, February 2015. Date last accessed 31 May 2018. www.quickheal.co.in/resources/threat-reports
8. Internet Security Threat Report (ISTR): Technical report, Symantec Corporation (2016). Date last accessed 31 May 2018
9. Sharma, A., Sahay, S.K.: An effective approach for classification of advanced malware with high accuracy. *Int. J. Secur. Appl.* **10**(4), 249–266 (2016)
10. Sahay, S.K., Sharma, A.: Grouping the executables to detect malware with high accuracy. *Procedia Comput. Sci.* **78**, 667–674 (2016)
11. Feizollah, A., Anuar, N.B., Salleh, R., Suarez-Tangil, G., Furnell, S.: Androdialysis: analysis of android intent effectiveness in malware detection. *Comput. Secur.* **65**, 121–134 (2017)
12. Das, S., Liu, Y., Zhang, W., Chandramohan, M.: Semantics-based online malware detection: towards efficient real-time protection against malware. *IEEE Trans. Inform. Forens. Secur.* **11**(2), 289–302 (2016)
13. Saracino, A., Sgandurra, D., Dini, G., Martinelli, F.: Madam: effective and efficient behavior-based android malware detection and prevention. *IEEE Trans. Depend. Sec. Comput.* **15**(1), 83–97 (2017)
14. Sharma, A., Sahay, S.K., Kumar, A.: Improving the detection accuracy of unknown malware by partitioning the executables in groups. In: Advanced Computing and Communication Technologies, pp. 421–431. Springer, Berlin (2016)
15. Sharma, A., Sahay, K, S.: An investigation of the classifiers to detect android malicious apps. In: Proceedings of ICICT Information and Communication Technology, vol. 625, pp. 207–217. Springer, Berlin (2017)
16. Sharma, A., Sahay, S.K.: Evolution and detection of polymorphic and metamorphic malwares: a survey. *Int. J. Comput. Appl.* **90**(2), 7–11 (2014)
17. Security software–Statistics & Facts: <https://www.statista.com/topics/2208/security-software> (2016). Date last accessed 21 Mar 2018

18. You, I., Yim, K.: Malware obfuscation techniques: a brief survey. In: 2010 International Conference on Broadband, Wireless Computing, Communication and Applications, Nov 2010, pp. 297–300
19. Stallings, W.: Network Security Essentials: Applications and Standards. Pearson Education India (2007)
20. Szor, P.: The Art of Computer Virus Research and Defense. Pearson Education (2005)
21. Beaucamps, P.: Advanced polymorphic techniques. *Int. J. Comput. Sci.* **2**(3), 194–205 (2007)
22. You, I., Yim, K.: Malware obfuscation techniques: a brief survey. In: 2010 International conference on Broadband, Wireless Computing, Communication and Applications, pp. 297–300. IEEE (2010)
23. Shah, A.: Approximate disassembly using dynamic programming. Ph.D. thesis, Citeseer (2010)
24. Rad, B.B., Masrom, M., Ibrahim, S.: Camouflage in malware: from encryption to metamorphism. *Int. J. Comput. Sci. Network Secur.* **12**(8), 74–83 (2012)
25. Rad, B.B., Masrom, M., Ibrahim, S.: Evolution of computer virus concealment and anti-virus techniques: a short survey. arXiv preprint [arXiv:1104.1070](https://arxiv.org/abs/1104.1070) (2011)
26. Szor, P., Ferrie, P.: Hunting for metamorphic. In: Virus Bulletin Conference, pp. 123–144 (2001)
27. Austin, T.H., Filiol, E., Josse, S., Stamp, M.: Exploring hidden markov models for virus analysis: a semantic approach. In: 46th Hawaii International Conference on System Sciences (HICSS), pp. 5039–5048. IEEE (2013)
28. Wong, W., Stamp, M.: Hunting for metamorphic engines. *J. Comput. Virol.* **2**(3), 211–229 (2006)
29. Griffin, K., Schneider, S., Hu, X., Chiueh, T.C.: Automatic generation of string signatures for malware detection. In: Recent Advances in Intrusion Detection, pp. 101–120. Springer, Berlin (2009)
30. Tran, N.P., Lee, M.: High performance string matching for security applications. In: 2013 International Conference on ICT for Smart Society (ICISS), pp. 1–5. IEEE (2013)
31. Ddcreateur: Antivirus 2004, [database on the internet] (March 2014), <http://files.codes-ources.com/fichierFullscreen.aspx?id=21418&f=virussignatur-es.txt&lang=en>
32. Grosso, N.D.: It's time to rethink your corporate malware strategy, 24 Feb, 2002. <https://www.sans.org/reading-room/whitepapers/malicious/its-time-rethink-corporate-malware-strategy-124>
33. Yanfang Ye, Tao Li, D.A.S.S.I.: A survey on malware detection using data mining techniques. *ACM Comput. Surv. (CSUR)* **50**(3), 41:1–41:40 (2017)
34. Inc, D.: 3% to 5% of enterprise assets are compromised by bot-driven targeted attack malware, 2 Mar 2009. <https://www.prnewswire.com/news-releases/3-to-5-of-enterprise-assets-are-compromised-by-bot-driven-targeted-attack-malware-61634867.html>
35. Harley, D., Lee, A.: Heuristic analysis—detecting unknown viruses. Technical report (2007). Date last accessed 31 May 2018
36. Mathur, K., Hiranwal, S.: A survey on techniques in detection and analyzing malware executables. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**(4), 422–428 (2013)
37. Govindaraju, A.: Exhaustive statistical analysis for detection of metamorphic malware (2010)
38. Xu, M., Wu, L., Qi, S., Xu, J., Zhang, H., Ren, Y., Zheng, N.: A similarity metric method of obfuscated malware using function-call graph. *J. Comput. Virol. Hack. Techniq.* **9**(1), 35–47 (2013)
39. Xu, J.Y., Sung, A.H., Chavez, P., Mukkamala, S.: Polymorphic malicious executable scanner by api sequence analysis. In: Fourth International Conference on Hybrid Intelligent Systems (HIS'04), pp. 378–383. IEEE (2004)
40. Mihai Christodorescu, Somesh Jha, S.A.S.D.S., Bryant, R.E.: Semantics-aware malware detection. In: Proceedings of the 2005 IEEE Symposium on Security and Privacy (SP '05), pp. 32–46. ACM, May 2005
41. Christodorescu, M., Kinder, J., Jha, S., Katzenbeisser, S., Veith, H.: Malware normalization. Technical report, University of Wisconsin (2005). <http://pages.cs.wisc.edu/~mihai/publications/Malware%20Normalization/Malware%20Normalization.pdf>

42. Armoun, S.E., Hashemi, S.: A general paradigm for normalizing metamorphic malwares. In: 2012 10th International Conference on Frontiers of Information Technology (FIT), pp. 348–353. IEEE (2012)
43. Mitchell, T.M.: Machine Learning. McGraw-Hill, Boston (1997)
44. Moskovich, R., Elovici, Y., Rokach, L.: Detection of unknown computer worms based on behavioral classification of the host. *Comput. Stat. Data Anal.* **52**(9), 4544–4566 (2008)
45. Alazab, M., Venkatraman, S., Watters, P., Alazab, M.: Zero-day malware detection based on supervised learning algorithms of API call signatures. In: Proceedings of the Ninth Australasian Data Mining Conference, vol. 121, pp. 171–182. Australian Computer Society, Inc. (2011)
46. Siddiqui, M., Wang, M.C., Lee, J.: A survey of data mining techniques for malware detection using file features. In: Proceedings of the 46th Annual Southeast Regional Conference on xx, pp. 509–510. ACM (2008)
47. Jacob, G., Debar, H., Filiol, E.: Behavioral detection of malware: from a survey towards an established taxonomy. *J. Comput. Virol.* **4**(3), 251–266 (2008)
48. Grosse, K., Papernot, N., Manoharan, P., Backes, M., McDaniel, P.: Adversarial Perturbations Against Deep Neural Networks for Malware Classification (2016). <https://arxiv.org/abs/1606.04435?context=cs>
49. Yin, W., Kann, K., Yu, M., Schutze, H.: Comparative Study of CNN and RNN for Natural Language Processing (2017). <https://arxiv.org/abs/1702.01923>

Preventing Forgeries by Securing Healthcare Data Using Blockchain Technology



V. Vetriselvi, Sridharan Pragatheeswaran, Varatharajan Thirunavukkarasu and Amaithi Rajan Arun

Abstract Decentralization has gained a lot of attention due to its application in diverse fields. It is pioneered largely by bitcoin, a blockchain technology, and a financial application of decentralization, which has impacted a lot on how financial transactions happen in a secure manner. The advantage of using this technology is that there is no central authority to rely on. Thus, a decentralized storage of medical records would allow forgeries on the records to be reduced. We propose a solution to avoid forgery in healthcare sector using blockchain. The blockchain network in the proposed system will time-stamp and store healthcare management data and its associated files in the network storage. The network is decentralized; thus, the data is inherently secure. Yet this approach may create a storage exploitation and may lead to breakdown of the system. However, a machine learning-based classification model is used to decide upon which records that get into the blockchain to reduce the required storage. Hence, a system to securely store healthcare data using blockchain technology can be implemented or created.

Keywords Decentralization · Blockchain · Healthcare · Medical records · Classification · Machine learning

V. Vetriselvi (✉) · S. Pragatheeswaran · V. Thirunavukkarasu · A. R. Arun
Department of Computer Science and Engineering, College of Engineering,
Guindy, Chennai, India
e-mail: vetri@annauniv.edu

S. Pragatheeswaran
e-mail: sripragathee29@gmail.com

V. Thirunavukkarasu
e-mail: vthirunavukkarasu96@gmail.com

A. R. Arun
e-mail: arunamaithirajan@gmail.com

1 Introduction

Blockchain technologies are really fascinating the researchers largely due to their success of the bitcoin cryptocurrency. The main reason for this fascination is that the applications which ran only through a centralized authority can now operate as a fully decentralized authority and achieve the same functionality with the same amount of certainty, with blockchain in that place [1]. Satoshi Nakamoto explained functionalities of blockchain and proof of work [2]. Figure 1 shows the structure of blockchain [3]. With this huge advancement, researchers are trying to implement blockchain in different scenarios rather than this cryptocurrency scenario. Some of the areas where researchers are trying to implement blockchain are finance [4, 5], healthcare to utilities, IoT field [6], music industry, real estate, and the government sectors.

Mizrahi [7] proposed how assets that can be uniquely identified by one or more identifiers can be registered in blockchain. This can be used to verify ownership of an asset and trace the transaction history. Any physical or digital property such as real estate, automobiles, laptops, and other valuables can potentially be registered in blockchain. The ownership and transaction history can be validated by anyone.

Mettler [8] projected how blockchain can prevent the counterfeiting of given drugs in the pharmaceutical industry. Estonia [9] is implementing a blockchain-based healthcare management record to store it in a hacker-proof manner. In that project, logs of access of healthcare data and audit data are stored in blockchain. All the users can see when the data is accessed, but the access log cannot be modified.

Nowadays, healthcare data is more important since it can be modified by higher authorities. Hence, the security of this healthcare data becomes questionable. If patients get affected by medications or treatments given that cannot be proved by them, it can be easily counterfeited. This happens because we completely rely on the centralized record management system maintained. So, we need to build a system which records the medications or treatments given in such a way that it is impossible to be counterfeited even by the higher authorities. Hence, the original treatments

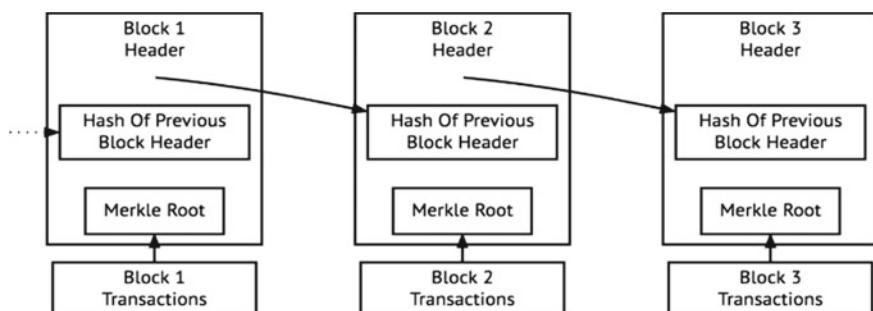


Fig. 1 Structure of blockchain

given cannot be hidden from the eye of society. To achieve this, we have to store all surgery details from patient's surgery approval date to surgery date in blockchain. Especially heart, brain, and cancer surgeries are major in India, which has to be monitored very seriously.

We are implementing blockchain for healthcare data. To reduce the storage in blockchain, we are storing only the patient's details who needs surgery by classifying their heartbeat signals.

Dokur and Olmez [10] developed a method for abnormal heartbeat sound decision-making process comprising three main stages. At the first stage, S1–S2 sounds are segmented [11], i.e., their timings are determined. S1–S2 sounds are used to extract the features of the heartbeat signals. At the second stage, feature vector elements are formed by using the wavelet plane. At the last stage, classification process is realized by a SVM learning algorithm [12].

Our system mainly focuses on securing the healthcare data using the blockchain technology. Since the healthcare data is very large in size, we have restricted ourselves to heart diseases. To fasten the process of diagnosis, CNN machine learning algorithm has been implemented to classify the heartbeat as normal or abnormal using the heartbeat signal recorded. The heartbeat signals have been first preprocessed. Then, noise has been removed by applying a low-pass filter. The features are extracted from the signal by FFT method. CNN model has been built using the features extracted, which is then later used for classifying heartbeat signals.

2 Proposed System

The machine learning phase classifies the heartbeat signals into normal or abnormal to find critical cases which have to be stored securely. We used blockchain to prevent those critical data from tampering. Additionally, we are analyzing the data stored in blockchain and identifying disease spread in the country with the help of GeoSmart contract.

2.1 Heartbeat Classification Phase

In this phase, we have implemented CNN to classify heartbeat signals into normal or abnormal. It includes training as well as testing phases. It consists of data preprocessing, feature extraction, and classification process.

Normalizer and Noise Remover. This module takes a normal WAV file as input. WAV file contains time domain signal values. There are some complex operations on these signals. We have applied random normalizing function which converts the amplitude values within –1 to 1 to make those operations easier. The heartbeat signals

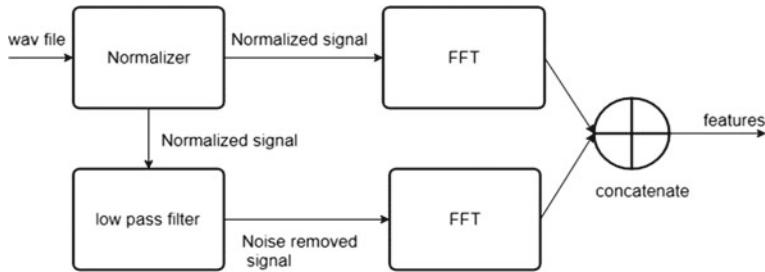


Fig. 2 Process of feature extraction

recorded might have extra noises which affect the accuracy of the classification. So, a Butterworth low-pass filter has been created to filter out frequencies greater than 2.5 Hz. However, normal human heartbeat frequencies maxes up to 2.5 Hz.

Feature Extraction. Features of each WAV file are needed to train the CNN classification algorithm. The coefficients of FFT have been taken as features. Formula of FFT is given in Eq. (1) [13].

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-j(2\pi|N|nk)} \quad (1)$$

where $X(k)$ is frequency domain samples, $x(n)$ is time domain samples, N is the size of FFT, and k varies from 0 to $N - 1$. To improve the accuracy of the classification, FFT has been done two times before and after noise removal and concatenate them. Flow of this procedure is given in Fig. 2.

CNN Classification Process. Training part of classification model has been done using convolutional neural network to get higher accuracy than other classification algorithms. For this CNN training, we have feature vectors as input. There are two hidden layers. We have two nodes in output layer to classify heartbeat signal as normal or abnormal. Initially, weights and biases are randomly assigned in this CNN; then weights and biases will be updated in upcoming iterations. Finally, this model predicts the heartbeat signal as normal or abnormal. A full classification process has been shown in Fig. 3.

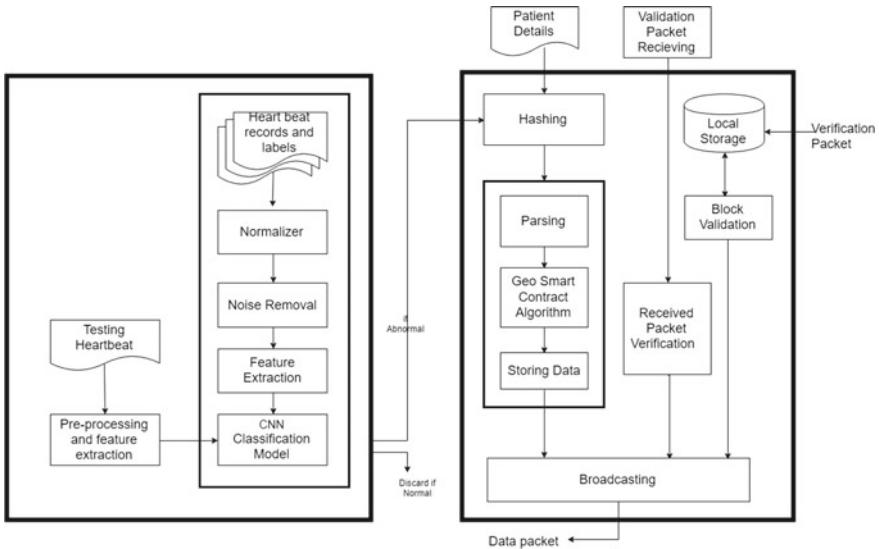


Fig. 3 Block diagram of a proposed system

2.2 Transaction Generation Phase

When the heartbeat signal is classified as abnormal, the patient has to undergo many screening processes. Once the surgery has been confirmed, the patient's details should be stored in blockchain. This phase comprises hashing the patient details, analysis of disease spread, and verification processes. Every transaction begins in this phase. Refer Fig. 3.

GeoSmart contract. Geo location data is extracted from patient details. This module will be triggered whenever a transaction is added to the blockchain. Here the flow of smart contract is given in Figure 4. Initially, smart contract parses the transaction, gets the location of the patient, and then alters the cumulative spread identity. GeoSmart contract gives us the survey of heart disease spread in all cities using geolocation data from the patient details.

3 Experimental Results

This section explains about the results obtained in CNN classification process and GeoSmart contract procedure which are described in previous sections.

3.1 Dataset Used

Heart signal recordings in the physionet heart signal database were sourced from several contributors, collected at either a clinical or nonclinical environment, from both healthy subjects and patients. The training set consists of five folders (A through E) containing a total of 3126 heart signal recordings, lasting from 5 to 120 s. The recordings were collected from different locations on the body. In both training and testing sets, heart signal recordings were divided into normal as well as abnormal heart signal recordings. The normal recordings were from healthy subjects, and the abnormal ones were from patients with a confirmed cardiac diagnosis. All recordings have been provided as .WAV format [14].

3.2 Accuracy in Classification of Heartbeat Signals

The cardiac sounds are selected from the physionet heart sound database. We used sixth-order Butterworth filter with cutoff frequency 2.5 Hz. All signals above 2.5 Hz are filtered out. Then we applied FFT algorithm on it and got features from that to train the CNN model. Accuracy of the prediction over the different size of dataset has been shown in Fig. 5.

When the size of dataset increases, the accuracy of the classification model also increases.

3.3 Comparison with Previous Approaches

We compare accuracy of our system with previous works with the help of following Eq. (2).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TF} + \text{TN} + \text{FN}} \quad (2)$$

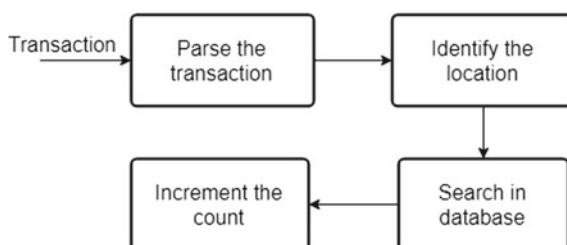


Fig. 4 Flow of GeoSmart contract

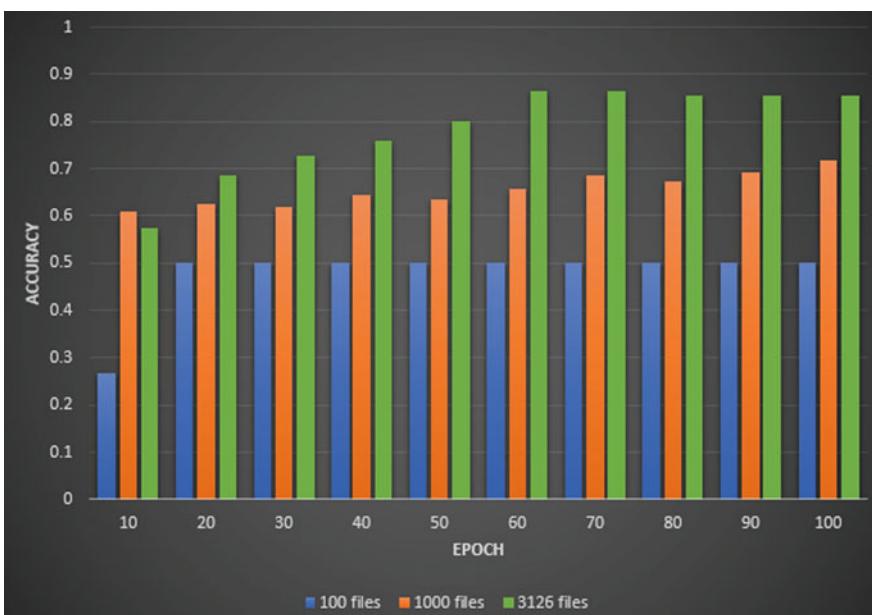
Table 1 Comparison with previous approaches

Approach	Accuracy
NN-2 layers [15]	0.79
DNN [16]	0.80
AdaBoost and CNN [17]	0.86
Our system	0.88

TP, FP, TN, and FN represent true positive, false positive, true negative, and false negative, respectively. Comparison with previous works with our work is shown in Table 1. Accuracy of our system is 88%.

3.4 Blockchain GeoSmart contract and Transaction Acknowledgement

Figure 6 states that the system is updated on execution of each transaction by GeoSmart contract. The bundled transaction is parsed, and the frequency of occurrence of surgeries is viewed in the front end.

**Fig. 5** Comparison between accuracy when varying size of dataset

State	Surgeries
State 1	23
State 2	45
State 3	90

Fig. 6 Result of GeoSmart contract

```
2018-04-24 08:05:44.594 UTC [shim] beforeInit -> DEBU 181 [78832eb8]Received INIT, initializing chaincode
2018-04-24 08:05:44.594 UTC [qscc] Init -> INFO 182 Init QSCC
2018-04-24 08:05:44.594 UTC [msp] GetLocalMSP -> DEBU 183 Returning existing local MSP
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 184 [78832eb8]Init get response status: 200
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 185 [78832eb8]Init succeeded. Sending COMPLETED
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 186 [78832eb8]Move state message COMPLETED
2018-04-24 08:05:44.594 UTC [shim] handleMessage -> DEBU 187 [78832eb8]Handling ChaincodeMessage of type: COMPLETED(state:ready)
2018-04-24 08:05:44.594 UTC [shim] func1 -> DEBU 188 [78832eb8]send state message COMPLETED
2018-04-24 08:05:44.594 UTC [chaincode] processStream -> DEBU 189 [78832eb8]Received message COMPLETED from shim
2018-04-24 08:05:44.594 UTC [chaincode] HandleMessage -> DEBU 190 [78832eb8]Fabric side Handling ChaincodeMessage of type: COMPLETED in state ready
2018-04-24 08:05:44.594 UTC [chaincode] HandleMessage -> DEBU 190 [78832eb8-dfb9-4e79-bbe5-b9179b2120a8]HandleMessage- COMPLETED. Notify
2018-04-24 08:05:44.594 UTC [chaincode] notify -> DEBU 191 notifying TxId:78832eb8-dfb9-4e79-bbe5-b9179b2120a8
2018-04-24 08:05:44.594 UTC [chaincode] Execute -> DEBU 19d Exit
```

Fig. 7 Internal log of peer (hyperledger fabric)

Figure 7 is the internal log of one of the peer nodes. The peer is installed with chaincode. Logs state that the peer has received a transaction for chaincode invocation and the transaction identity (Txid) of executed transaction is notified. Thus, we can infer that the fabric is storing transaction in the blockchain storage.

4 Conclusion and Future Works

In our system, CNN has been used for the classification of heartbeat signals and hyperledger has been used for building blockchain. If the heartbeat is classified abnormal, then the corresponding patient has to undergo a surgery if needed. The surgery details have been considered as a transaction and broadcast in the blockchain network. Hence, the surgery details cannot be modified. In future, any machine learning algorithm which gives more efficient results than CNN can be used. X-ray reports, blood reports, etc., are not being stored in blockchain; these important reports also can be stored in blockchain. Due to scalability issues, we have restricted ourselves to heart surgeries; this can be extended to support all kinds of treatments and surgeries.

References

1. Owenson, G., Dennis, R., Aziz, B.: A temporal blockchain: a formal analysis. In: International Conference on Collaboration Technologies and Systems, pp. 430–437 (2016)

2. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. <https://bitcoin.org/bitcoin.pdf> (2008)
3. Bitcoin: Bitcoin developer guide. <https://bitcoin.org/en/developer-guide#block-chain> (2017)
4. Pattanayak, P., Verma, S., Crosby, M., Nachiappan, Kalyanaraman, V.: Blockchain technology: beyond bitcoin. <http://scet.berkeley.edu/wpcontent/uploads/AIR-2016-Blockchain.pdf> (2016)
5. Tschorsh, F., Scheuermann, B.: Bitcoin and beyond: a technical survey on decentralized digital currencies. *IEEE Commun. Surv. Tutor.* (2015)
6. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. *IEEE Access J. Rapid Open Access Publ.* **04**, 2292–2303 (2016)
7. Mizrahi, A.: A blockchain based property ownership recording system. <https://chromaway.com/papers/A-blockchainbased-property-registry.pdf>
8. Mettler, M.: Blockchain technology in healthcare the revolution starts here. In: IEEE 18th International Conference on e-Health Networking, Applications and Services (2017)
9. Estonian citizens will soon have the world's most hack proof health-care records. <https://qz.com/628889/thiseastern-european-country-is-moving-its-health-records-to-theblockchain/>
10. Dokur, Z., Olmez, T.: Heart sound classification using wavelet transform and incremental self-organizing map. *Digital Signal Proc.* **18**, 951–959 (2008). Elsevier
11. Shervegar, M.V., Bhat, G.V.: Automatic segmentation of phonocardiogram using the occurrence of the cardiac events. *Inf. Med. Unlock. J.* **9**, 6–10 (2017). Elsevier
12. Azman, A., Jantan, A., Safara, F., Doraisamy, S., Ranga, A., Ramaiyah, A.: Multi-level basis selection of wavelet packet decomposition tree for heart sound classification. *Comput. Biol. Med.* **43**, 1407–1414 (2013). Elsevier
13. Wada, T.: 64 point fast fourier transform circuit. <http://www.ie.u-ryukyu.ac.jp/wada/design07/specs.html> (2006)
14. <http://physionet.org/physiobank/database/challenge/2016/>. Accessed date: 29 Dec 2017
15. Grzegorczyk, I., Soliński, M., Łepeł, M., Perka, A., Rosiński, J., Rymko, J., Stępień, K., Gierałtowski, J.: PCG classification using a neural network approach. In: 2016 Computing in Cardiology Conference (CinC), pp. 621–624. IEEE (2016)
16. Nassralla, M., El Zein, Z., Hajj, H.: Classification of normal and abnormal heart sounds. In: 2017 Fourth International Conference on Advances in Biomedical Engineering (ICABME), pp. 1–4. IEEE (2017)
17. Potes, C., Parvaneh, S., Rahman, A., Conroy, B.: Ensemble of feature based and deep learning-based classifiers for detection of abnormal heart sounds. In: 2016 Computing in Cardiology Conference (CinC), pp. 621–624. IEEE (2016)

A Survey of Deep Learning Techniques for Medical Diagnosis



Abdul Mueed Hafiz and Ghulam Mohiuddin Bhat

Abstract With the advent of new technologies in artificial intelligence and machine learning, the medical community has taken a strong notice of the potential of these technologies for addressing automation. Deep learning is one of these technologies which has been chosen by the research community for advancing its medical applications. This survey paper serves the research community twofold. First, it gives researchers an introduction to the basic technologies involved in deep learning. Second, it gives the readers insight into the state of the art in the field of medical applications of deep learning, particularly for medical imaging technologies.

Keywords Deep learning · Medical imaging · Artificial intelligence

1 Introduction

Deep learning [1] is gathering significant momentum in data analysis and has been amongst top ten new technologies for the year 2013 [2]. Deep learning is an enhanced version of neural nets having many layers in turn allowing higher abstraction. This further leads to improvement in predictions from the data available [1]. As of now, deep learning has emerged as the frontier artificial intelligence tool of choice for most of the machine-vision research areas. Deep learning has experienced substantial interest from the medical machine-vision community [3–6]. This is because deep learning is able to entertain huge databases of medical imaging, is able to perform data analysis across multiple modes and is able to transfer learning of the neural net across different datasets. As an example, in magnetic resonance imaging (MRI), the applications of deep learning to its images [7] include auto-segmentation [8–18], auto-upgradation without upgrading the MRI device [19], MRI to computerized tomography (CT) scan transformation [20], MRI to positron emission tomog-

A. M. Hafiz (✉) · G. M. Bhat

Department of Electrical and Electronics Engineering, Institute of Technology, The University of Kashmir, Srinagar 190006 , J&K, India
e-mail: mueedhafiz@uok.edu.in

raphy (PET) scan transformation [21], etc. For an introduction to deep learning, one may refer to [1].

2 Overview of Deep Learning

Deep learning [22] reflects recent advances in artificial neural network (ANN) programming, and it represents a technique to train deep neural networks. Essentially, any ANN with more than two layers is deep. Hinton in 1984 was the first scientist to train deep nets. Earlier in 1943, Pitts had introduced the multilayer perceptron. Deep learning typically comprises features like partially labelled data, rectified linear units (ReLUs), convolutional neural networks (CNNs) and dropout. Each of these components is discussed in the following sections.

2.1 *Partially Labelled Data*

The term ‘partially labelled data’ refers to training data that has missing group or class labels. Some deep nets are able to train on partially labelled data. Their neural net weights are initialized without assigning to them their outputs, etc. Their different layers can be trained without using the class labels. Once this unsupervised training has taken place in the convolutional net, a supervised training is used to tweak them.

2.2 *Rectified Linear Units*

The rectified linear unit, or ReLU, is the main activation function for hidden layer neurons of deep nets. For the deep belief neural net (DBNN), the restricted Boltzmann machine (RBM) is common. Besides ReLU activation function, deep nets also use the linear/softmax activation function in their output layer units. This function is used for networks which are used for regression, classification, etc. The ReLU function can be mathematically expressed as:

$$\Phi(x) = \max(0, x) \quad (1)$$

2.3 *Convolutional Neural Networks*

Convolution is a mathematical process which is frequently used in deep learning. Convolution permits computer-vision nets to behave almost like biological systems

and with higher accuracy. A technique used here is assigning sparse links amongst two different layers wherein every weight is not indicated. Such layers are called ‘dense layers’, and the same process is called sparse connectivity. Besides not having weights for every link, convolutional nets share connection/link weights. LeNET-5 neural net architecture [23] is an implementation of a convolutional net. The model is usually used for image classification. This architecture is the same as a feed-forward neural network in many ways. Data flows through the network. However, in contrast to conventional nets, LeNET-5 contains different layers.

3 Tools for Deep Learning

Many tools/software have been developed for training graphical processing units (GPUs) used in deep learning neural nets. H₂O is a machine learning software framework in Java used in the Web as a service. H₂O does support various programming languages. H₂O is compatible with Java, Scala, Python, R and various languages which can interface with its API. Theano is a Python-based software. Although it avoids direct implementation of deep neural networks, it comprehensively provides many mathematical tools for creation of deep neural net apps. Due to the fact that Theano’s support for deep learning is missing, many software have been integrated into Theano for the purpose of making deep learning implementation easy. Lasagne and Nolearn are a package pair used together, for the same. ConvNetJS is a JavaScript software which provides implementation of many deep learning techniques, especially in the area of convolutional nets. ConvNetJS aims at creation of examples of deep learning on the Internet.

4 Medical Applications of Deep Learning

4.1 Segmentation of Medical Images

Medical image segmentation of body parts and their components, e.g. blood cells, leads to volumetric and shape-related clinical analysis of tumours, etc. [24]. Segmentation is also an important first precursor to computer-aided detection/classification. Segmentation in medical image analysis is defined as the process of detecting of the voxel sets contributing to formation of an object’s contour. It is a deeply researched medical application of deep learning techniques [7, 8]. Segmentation applications/algorithms vary widely in methodology, e.g. use of unique CNN-based architectures, combinations of feature maps, skip connections, etc. Further, the segmentation may be 2D or 3D (e.g. in a stack of MRI slices, ultrasound video, etc.).

In [25], the authors present a pooling technique applicable to CNNs for classification of HEp-2 cells. This technique known as *K*-support spatial pooling is a

spatial-pooling-based technique in convolutional nets wherein besides integration of largest activated magnitude, account is taken of response of relative activated features of a specific net unit. The convolutional nets with the above pooling technique are used for classification of HEp-2 cells. Versatile results have been obtained. In [26], authors propose a CNN-based cell segmentation technique used in microscopy image analysis. The approach proceeds as follows. In the first step, this technique produces ROI indicators with the help of a novel modification of the fully convolutional net also known as FCN [27]. Afterwards, another convolutional net produces image-segmentation masks used for every ROI indicator. The technique has reportedly outperformed three versatile contemporary techniques used for cell-image segmentation [28–30] applied to three databases [29, 31, 32] for different appearances and lightings.

Excellent segmentation of lymph nodes has been obtained using a 2.5-dimensional convolutional net in which the three-dimensional object-of-interest is broken down into its views of coronal, axial and sagittal sections. The 2.5-dimensional convolutional net, in comparison to the three-dimensional convolutional net, alleviates both the processing overload and the dimension-related issues [33]. 2D and 3D convolutional nets [34, 35] have also been used to segment multiple sclerosis (MS) lesions. In [11], the authors propose a longitudinal multiple-view convolutional net to segment these lesions. The input to the convolutional nets used is from different image patches. This novel convolutional net used and conveniently uses the longitudinally segmented data. This segmentation technique was used on the ISBI 2015 database, and it gave results at par and even better than those of manual segmentation. In [10], the authors use a FCN for liver segmentation and subsequent lesion detection. The convolutional net model is derived from the sixteen layered Visual Geometry Group Network (VGG-16) [36] and is a novel modification of the same. The model gives a heat-map which is subsequently used for detecting lesions.

In [18], a new technique is given which segments stacks of CT scans using an FCN, which gives per-voxel multiple class classification. The FCN model used is based on a VGG-16 network [36]. The model is a full convolutional net with both convolution and deconvolution components. The technique achieved segmentation of nineteen classes of organs in the torso (all major organs). The results of the experiments, when the proposed technique was used, were better than when manual intervention was present.

Recurrent neural networks (RNNs) have also become popular for segmentation tasks in medical image analysis. In [37], the authors report an RNN which segments the images of H&E histopathology. In [38], the authors use a novel 3D long short-term memory recurrent neural network (LSTM-RNN) having six-dimensional convolution. In [39], a 3D RNN with gated recurrent units (GRUs) [40] has been used to segment MRI brain images. In [41], the authors have fused a two-dimensional network [42] with GRUs for segmenting three-dimensional voxels.

One problem with techniques used to classify voxels is spurious output. To overcome the problem, some researchers have tried fusion of FCNs with Markov random fields (MRFs), e.g. in [43, 44], and conditional random fields (CRFs), respectively, e.g. in [45–50], for the purpose of refinement of the segmented regions. Mostly,

MRFs/CRFs are fused at the start of the convolutional nets which in turn act by label regularization.

4.2 Classification/Detection

Classification is one the original applications of artificial intelligence (AI) and machine learning. CNNs have shown robustness and excellent results in the field of object classification in machine vision [24]. A relevant technique in this regard, viz. transfer learning [51, 52], refers to the use of a pre-trained CNN for classification of a new image dataset in hand. In [53], the authors investigate transferring a ImageNet-trained CNN to detect kidneys in ultrasound images. They also investigate the extent of transfer which affects detection performance. The authors experimentally demonstrate that their CNN outperforms contemporary versatile techniques. The authors further report that a hybrid of their technique and another reported state-of-the-art technique gives better accuracy by a margin of 20%. In [54], the authors have introduced a mammogram image tumour detection system that combines segmentation modules related to the breast, as well as fibroglandular tissues, into a region-based convolutional network (R-CNN) [55]. Experimental outcomes demonstrate the robustness and accuracy of the proposed network.

In [56], a novel technique is presented, using a deep CNN which is a modification of AlexNet [57] for blood vessel detection in ultrasound images. Caffe framework [58] has been used for the classifier. The FAST medical image computing framework [59] has been used for the vessel candidate search implementation. This novel technique is able to localize in real time the vessels in the B-mode ultrasound images giving a classification accuracy of 94.5% which is a much better than classification accuracy of previous techniques viz. 84% for a common database. Sleep apnoea detection has also been studied elsewhere [60].

4.3 Image Reconstruction

Image reconstruction here refers to upgradation of resolution of images from related devices like MRIs, e.g. 7 T (7 T) from 3 T (3 T) [24]. Convolutional nets have also successfully been used for generation of high-resolution images from low-resolution images by generating a nonlinear mapping using paired low- and high-resolution images during training.

3 T to 7 T MRI

In [61], a resolution enhancement method is presented, which proceeds by learning a map across low- and high-resolution images. In [62], a compressive sensing-based technique is presented for increasing the resolution of MRI scans. In [19], a deep convolutional net is presented, which utilizes various feature maps for upgrading

resolution of MRI images. The convolutional net is trained patch-wise for a 3 T image, on features related to its appearance and anatomy. The convolutional net outputs the central voxel of the corresponding 7 T image patch. Next, 3 T–7 T image mapping is done by the convolutional net. After technical evaluation, the results show that the proposed technique outperforms other versatile contemporary techniques used for comparison.

MRI to CT

In [20], the authors propose a technique for mapping images pertaining to different devices, e.g. from MRI to CT. The mapping is done by a three-dimensional FCN. The process proceeds as follows. First, the input image is broken into overlapping segments. For each segment, the neural net predicts its corresponding device segment. In the end, the generated output segments are fused into an image by taking the averages of the greyscales present in the segments with overlap. The technique has been tested on a database of pelvic CT and MRI images. The experimental results show that proposed technique is effective as it outperforms other contemporary versatile techniques.

MRI to PET

In [21], the authors have proposed a technique wherein they reconstruct PET images from MRI images by a 3D convolutional net. The technique was evaluated using Alzheimer's disease neuroimaging initiative (ADNI) database and has shown good results.

4.4 Denoising/Normalization

Image enhancement applications like denoising and intensity normalization have been researched limitedly in deep learning [24]. In [63], sparse auto-encoders (SAEs) have been used to normalize images of histopathology stained by H&E. [64] used CNNs to perform denoising in dynamic contrast-enhanced MRI (DCE-MRI) time-series. The technique uses expert deep neural nets, where each of these is trained on various noise features and other curves of different prototypes. Finally, the best mapping is chosen for classification.

5 Conclusion

Deep learning is gathering momentum as the machine learning tool of choice for computer vision. It feeds on large amounts of hardware and data. Though immensely successful, its biological evidence remains a question mark, and this leads to the possibility of new avenues of machine learning which are unfolding as we speak, ever since after the success of nature. Other possible avenues for future work vis-à-vis deep learning medical applications are context-aware deep learning techniques;

fusion of statistical techniques with deep learning architectures, training procedures and classification procedures; use of multidimensional data by deep learning, etc.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
2. MIT Technology Review identifies the 10 most important technology milestones of the past year, M.I.T. Technology Review (23 April, 2013). online: <https://www.technologyreview.com/s/513981/the-10-breakthroughtechnologies-of-2013/>
3. Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: overview, challenges and the future. In: Classification in BioApps, pp. 323–350. Springer (2018)
4. Selvathi, D., Poornila, A.A.: Deep learning techniques for breast cancer detection using medical image analysis. In: Biologically Rationalized Computing Techniques for Image Processing Applications, pp. 159–186. Springer (2018)
5. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. arXiv preprint [arXiv:1702.05747](https://arxiv.org/abs/1702.05747) (2017)
6. Greenspan, H., van Ginneken, B., Summers, R.M.: Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging* **35**(5), 1153–1159 (2016)
7. Liu, J., Pan, Y., Li, M., Chen, Z., Tang, L., Lu, C., Wang, J.: Applications of deep learning to MRI images: a survey. *Big Data Mining Anal.* **1**(1), 1–18 (2018)
8. Yoo, Y., Tang, L.Y., Brosch, T., Li, D.K., Kolind, S., Vavasour, I., Rauscher, A., MacKay, A.L., Traboulsee, A., Tam, R.C.: Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *NeuroImage Clin.* **17**, 169–178 (2018)
9. Avendi, M., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **30**, 108–119 (2016)
10. Ben-Cohen, A., Diamant, I., Klang, E., Amitai, M., Greenspan, H.: Fully convolutional network for liver segmentation and lesions detection. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 77–85. Springer (2016)
11. Birenbaum, A., Greenspan, H.: Longitudinal multiple sclerosis lesion segmentation using multi-view convolutional neural networks. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 58–67. Springer (2016)
12. de Brébisson, A., Montana, G.: Deep neural networks for anatomical brain segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 20–28 (2015)
13. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., Pal, C.: The importance of skip connections in biomedical image segmentation. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 179–187. Springer (2016)
14. Hawaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H.: Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017)
15. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, pp. 131–143. Springer (2015)
16. Pereira, S., Pinto, A., Alves, V., Silva, C.A.: Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **35**(5), 1240–1251 (2016)
17. Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage* **108**, 214–224 (2015)

18. Zhou, X., Ito, T., Takayama, R., Wang, S., Hara, T., Fujita, H.: Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 111–120. Springer (2016)
19. Bahrami, K., Shi, F., Rekik, I., Shen, D.: Convolutional neural network for reconstruction of 7 T-like images from 3 T MRI using appearance and anatomical features. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 39–47. Springer (2016)
20. Nie, D., Cao, X., Gao, Y., Wang, L., Shen, D.: Estimating CT image from MRI data using 3D fully convolutional networks. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 170–178. Springer (2016)
21. Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S.: Deep learning based imaging data completion for improved brain disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 305–312. Springer (2014)
22. Heaton, J.: Artificial Intelligence for Humans, Volume 3: Deep Learning and Neural Networks. Heaton Research, Inc., St Louis, MO, USA (2015)
23. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
24. Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z.: Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings, vol. 10008. Springer (2016)
25. Han, X.-H., Lei, J., Chen, Y.-W.: HEp-2 cell classification using K -support spatial pooling in deep CNNs. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 3–11. Springer (2016)
26. Akram, S.U., Kannala, J., Eklund, L., Heikkilä, J.: Cell segmentation proposal network for microscopy image analysis. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 21–29. Springer (2016)
27. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(4), 640–651 (2017)
28. Akram, S.U., Kannala, J., Eklund, L., Heikkilä, J.: Cell proposal network for microscopy image analysis. In: 2016 IEEE International Conference on Image Processing (ICIP), pp. 3199–3203. IEEE (2016)
29. Arteta, C., Lempitsky, V., Noble, J.A., Zisserman, A.: Learning to detect cells using non-overlapping extremal regions. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 348–356. Springer (2012)
30. Magnusson, K.E., Jaldén, J.: A batch algorithm using iterative application of the Viterbi algorithm to track cells and construct cell lineages. In: 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), pp. 382–385. IEEE (2012)
31. Kainz, P., Urschler, M., Schulter, S., Wohlhart, P., Lepetit, V.: You should use regression to detect cells. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 276–283. Springer (2015)
32. Maška, M., Ulman, V., Svoboda, D., Matula, P., Matula, P., Ederra, C., Urbiola, A., España, T., Venkatesan, S., Balak, D.M.: A benchmark for comparison of cell tracking algorithms. Bioinformatics **30**(11), 1609–1617 (2014)
33. Roth, H.R., Lu, L., Seff, A., Cherry, K.M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., Summers, R.M.: A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 520–527. Springer (2014)
34. Ghafoorian, M., Platel, B.: Convolutional neural networks for MS lesion segmentation, method description of DIAG team. In: Proceedings of the 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge, pp. 1–2 (2015)

35. Vaidya, S., Chunduru, A., Muthuganapathy, R., Krishnamurthi, G.: Longitudinal multiple sclerosis lesion segmentation using 3D convolutional neural networks. In: Proceedings of the 2015 Longitudinal Multiple Sclerosis Lesion Segmentation Challenge, pp. 1–2 (2015)
36. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
37. Xie, Y., Zhang, Z., Sapkota, M., Yang, L.: Spatial clockwork recurrent neural network for muscle perimysium segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 185–193. Springer (2016)
38. Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J.: Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation. In: Advances in Neural Information Processing Systems, pp. 2998–3006 (2015)
39. Andermatt, S., Pezold, S., Cattin, P.: Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 142–151. Springer (2016)
40. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078) (2014)
41. Poudel, R.P., Lamata, P., Montana, G.: Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. In: International Workshop on Reconstruction and Analysis of Moving Body Organs, pp. 83–94. Springer (2016)
42. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
43. Shakeri, M., Tsogkas, S., Ferrante, E., Lippe, S., Kadoury, S., Paragios, N., Kokkinos, I.: Sub-cortical brain structure segmentation using F-CNN's. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 269–272. IEEE (2016)
44. Song, Y., Zhang, L., Chen, S., Ni, D., Lei, B., Wang, T.: Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning. IEEE Trans. Biomed. Eng. **62**(10), 2421–2433 (2015)
45. Alansary, A., Kamnitsas, K., Davidson, A., Khlebnikov, R., Rajchl, M., Malamateniou, C., Rutherford, M., Hajnal, J.V., Glocker, B., Rueckert, D.: Fast fully automatic segmentation of the human placenta from motion corrupted MRI. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 589–597. Springer (2016)
46. Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 442–450. Springer (2016)
47. Christ, P.F., Elshaer, M.E.A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., Rempfler, M., Armbrester, M., Hofmann, F., D'Anastasi, M.: Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 415–423. Springer (2016)
48. Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., Mok, V.C., Shi, L., Heng, P.-A.: Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans. Med. Imaging **35**(5), 1182–1195 (2016)
49. Fu, H., Xu, Y., Lin, S., Wong, D.W.K., Liu, J.: Deepvessel: retinal vessel segmentation via deep learning and conditional random field. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 132–139. Springer (2016)
50. Gao, M., Xu, Z., Lu, L., Wu, A., Nogues, I., Summers, R.M., Mollura, D.J.: Segmentation label propagation using deep convolutional neural networks and dense conditional random field. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1265–1268. IEEE (2016)
51. Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M.: Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans. Med. Imaging **35**(5), 1285–1298 (2016)

52. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
53. Ravishankar, H., Sudhakar, P., Venkataramani, R., Thiruvenkadam, S., Annangi, P., Babu, N., Vaidya, V.: Understanding the mechanisms of deep transfer learning for medical images. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 188–196. Springer (2016)
54. Akselrod-Ballin, A., Karlinsky, L., Alpert, S., Hasoul, S., Ben-Ari, R., Barkan, E.: A region based convolutional network for tumor detection and classification in breast mammography. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 197–205. Springer (2016)
55. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
56. Smistad, E., Løvstakken, L.: Vessel detection in ultrasound images using deep convolutional neural networks. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 30–38. Springer (2016)
57. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
58. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
59. Smistad, E., Bozorgi, M., Lindseth, F.: FAST: framework for heterogeneous medical image computing and visualization. *Int. J. Comput. Assist. Radiol. Surg.* **10**(11), 1811–1822 (2015)
60. Dey, D., Chaudhuri, S., Munshi, S.: Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomed. Eng. Lett.* **8**(1), 95–100 (2018)
61. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(2), 295–307 (2016)
62. Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., Ashok, A.: Reconnet: non-iterative reconstruction of images from compressively sensed measurements. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 449–458 (2016)
63. Janowczyk, A., Basavanhally, A., Madabhushi, A.: Stain normalization using sparse autoencoders (StaNoSA): application to digital pathology. *Comput. Med. Imaging Graph.* **57**, 50–61 (2017)
64. Benou, A., Veksler, R., Friedman, A., Raviv, T.R.: De-noising of contrast-enhanced MRI sequences by an ensemble of expert deep neural networks. In: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 95–110. Springer (2016)

Integrated Use of AHP and GIS Techniques for Selection of Artificial Water Recharge Sites



Shaikh Husen, Santosh Khamitkar, Parag Bhalchandra, Preetam Tamsekar, Govind Kulkarni and Kailas Hambarde

Abstract Water is one of the indispensable normal assets for the survival of life on the planet Earth. Interest for water is expanding step by step, with the consistently expanding populace, came about extreme water emergency. There is thus immense need for escalation of artificial water recharge sites for effective water management. This paper summarizes use of analytic hierarchy process (AHP) and geographic information system (GIS) in terms of multi-criteria decision support system in Indian scenario for selection of artificial water recharge sites. The framework was designed, formed and run over selected area. Six different thematic layers including geomorphology, soil, slope, LULC, lineament density, and drainage density were considered for study purpose. The Saaty's scale was used and suitable weights of selected layers with their features were assigned to generate eigenvectors, and finally, maps were integrated with weighted overlay analysis method of GIS.

Keywords Geographic information system (GIS) · Analytic hierarchy process (AHP) · Decision support system (DSS) · Remote sensing (RS)

S. Husen (✉) · S. Khamitkar · P. Bhalchandra · P. Tamsekar · G. Kulkarni · K. Hambarde
School of Computational Sciences, S.R.T.M. University, Nanded 431606, MS, India
e-mail: husen09@gmail.com

S. Khamitkar
e-mail: s_khamitkar@yahoo.com

P. Bhalchandra
e-mail: srtmun.parag@gmail.com

P. Tamsekar
e-mail: pritam.tamsekar@gmail.com

G. Kulkarni
e-mail: govindcoolkarni@gmail.com

K. Hambarde
e-mail: kailas.srt@gmail.com

1 Introduction

Groundwater is one of the essential commodity for human survival. It is available in limited quantity. Besides drinking, it is actively used in agricultural, municipal, and industrial developments. Its demand has increased over years, but it seems that the demand will never cease. Rainfall is a prime source of new water in India. Lots of rainwater is gone due to runoff, which is a problem arose due to poor rainwater harvesting practices or awareness. To overcome it, there are necessity groundwater recharge systems structures, such as *farm ponds, check dams, percolation tanks, and nala bunds* [1]. We can traditionally find out recharge sites to do so. Similarly, the remote sensing and geographic information system (GIS) tools can contribute to it more scientifically. Utilization of GIS and RS can likewise be considered for multi-criteria examination in resource assessment and hydro-geo-morphological mapping for water resource management. Analysis of RS data with the Survey of India (SOI) topographical sheets and secondary data with essential ground reality confirmation help in producing the benchmark data for groundwater focusing on. To find out the presence of groundwater sites using RS data and terrain features like soil, geomorphologic, lineament and hydrological characters. Arrangements of weights for the diverse themes are chosen in light of individual judgment, considering their virtual importance from the counterfeit revives perspective. These thematic layers can be coordinated GIS framework to distinguish appropriate sites for artificial recharge [2].

2 Groundwater Scenario in India and Study Area

The general commitment of precipitation to the nation's yearly groundwater asset is 68%, and the commitment of different assets, for example, canal leakage, return flow form irrigation, recharge from ponds, tanks, and water conservation structures taken together, is 32%. On account of the extending populace in the country, the national per capita yearly ease of use of water has decreased from 1816 m³ in 2001 to 1544 m³ in 2011 [3]. The groundwater level information for January 2016 demonstrates that out of the aggregate 14,974 wells examined, 1371 (9%) wells are indicating depth of water under 2 m bgl (meters below ground level), 4958 (33%) wells are indicating depth of water in the profundity scope of 2–5 m bgl, 5342 (36%) wells are indicating depth of water in the profundity scope of 5–10 m bgl, 2498 (17%) wells are demonstrating depth of water 1 in the profundity scope of 10–20 m bgl, 607 (4%) wells are indicating depth of water in the profundity scope of 20–40 m bgl, and the staying 198 (1%) wells are indicating depth of water more than 40 m bgl [4]. **The study area is Kalmnoori taluka of Hingoli district (Maharashtra state, India)** is taken as study area which lies within coordinate of latitudes between 19.670 north latitude and 77.330 east longitudes. Its area is about 945 km². It has an average elevation of 480 m (1574 feet). The *Kayadhu* is the main river which flows from study area.

The selected study area is extracted from the top sheet collected from the certified government organization, and it is georeferenced.

3 Research Methodology

In order to generate the decision support system for identification of artificial water recharge sites, a specific process is undertaken from similar work [5].

3.1 Spatial Data Collection

The required RS data of *CartoDEM* of 2.5 m resolution and 3.5 m resolution and satellite data of IRS P5 LISS-III, which are taken from the National Remote Sensing's Bhuvan portal [6]. Slope and drainage map extracted from *CartoDEM* whereas LULC extracted from LISS-III. Geomorphology and Lineament data in scale 1:50,000 are extracted from WMS layer of *Bhuvan portal* [7]. The details of hydrological soil group map are taken from *SWAT Indian* dataset which is available in www.swat.tamu.edu website [8].

3.2 Generation of Criteria Map Using GIS

For artificial water recharge sites, the formulation of criteria is done by using following six scientific parameters,

- Geomorphology map:** The geomorphology map is shown in Fig. 1. Fairly close to 88% of the total area having denudational origin, and 9.77% of the area is structural origin. Anthropogenic origin and water bodies covered 0.18 and 1.26%, respectively.
- II. **LULC:** The land use of the present area is shown in Fig. 2. Approximately 73% of the total area having cropland, 12.5% of the area is under scrubland and 6.5% of having rivers. Forest land and water bodies have (lake, dam) covered area of 4 and 3.5%, respectively.
- III. **Lineament density map:** The lineament has taken from *Bhuvan portal*, and the lineament density map is shown in Fig. 3.
- IV. **Slope map:** The slope map of the present study area is shown in Fig. 4, which reveals that most area of the southern part is found under high slope, whereas a small portion of the area comes under low slope.
- V. **Drainage density map:** The drainage density of the present area is distributed into five classes as per Table 1, and the reclassified drainage density map is shown in Fig. 5.

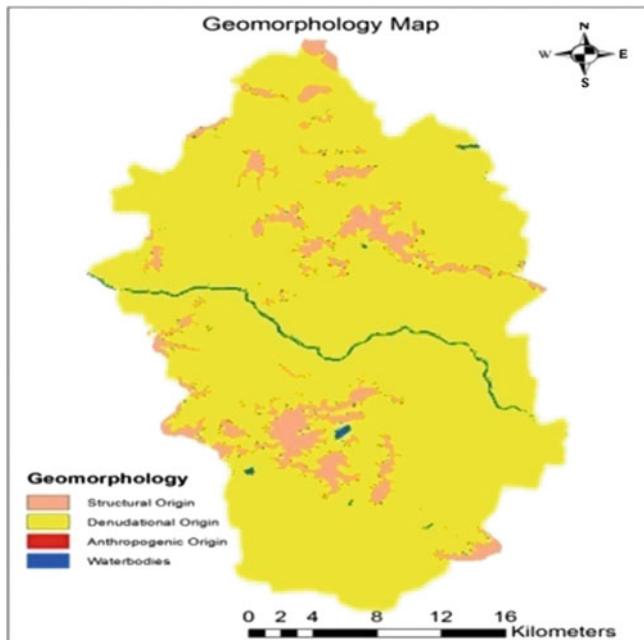


Fig. 1 Map for geomorphology

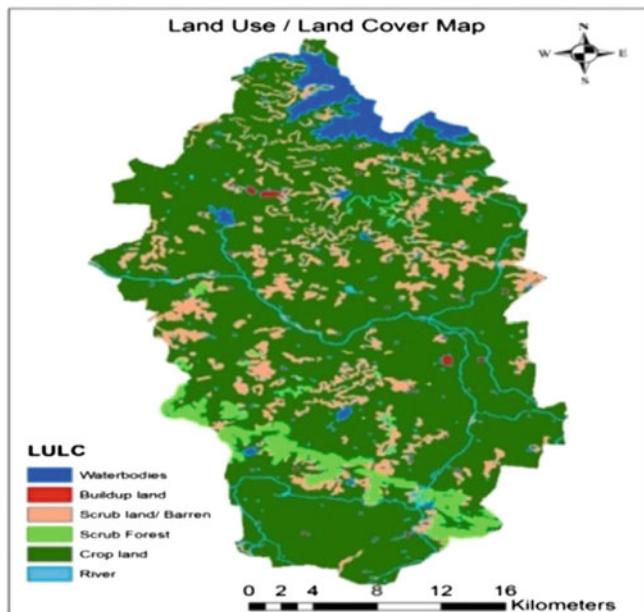


Fig. 2 Map for LULC

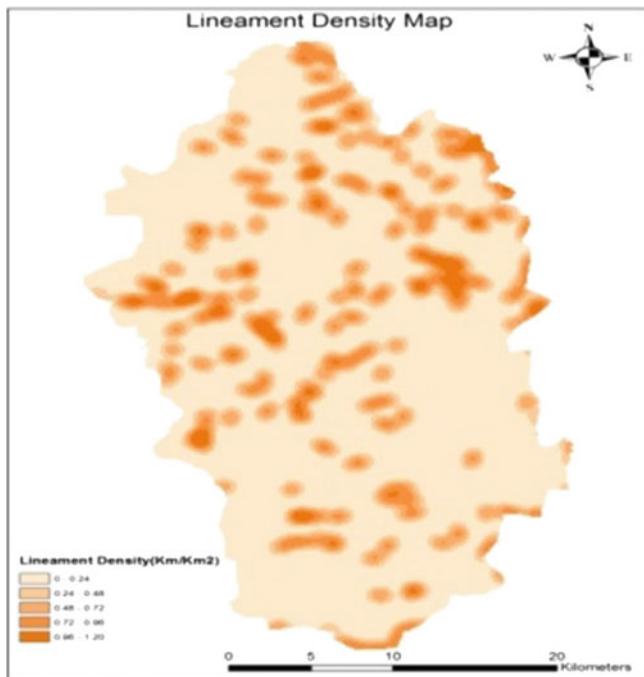


Fig. 3 Lineament density map

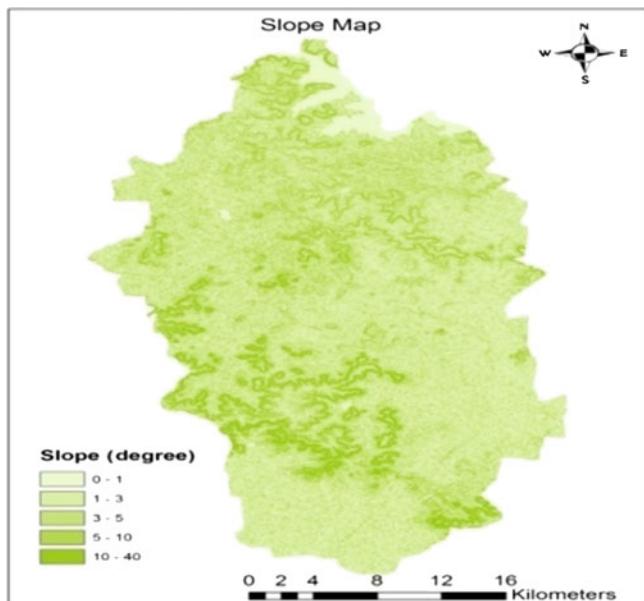
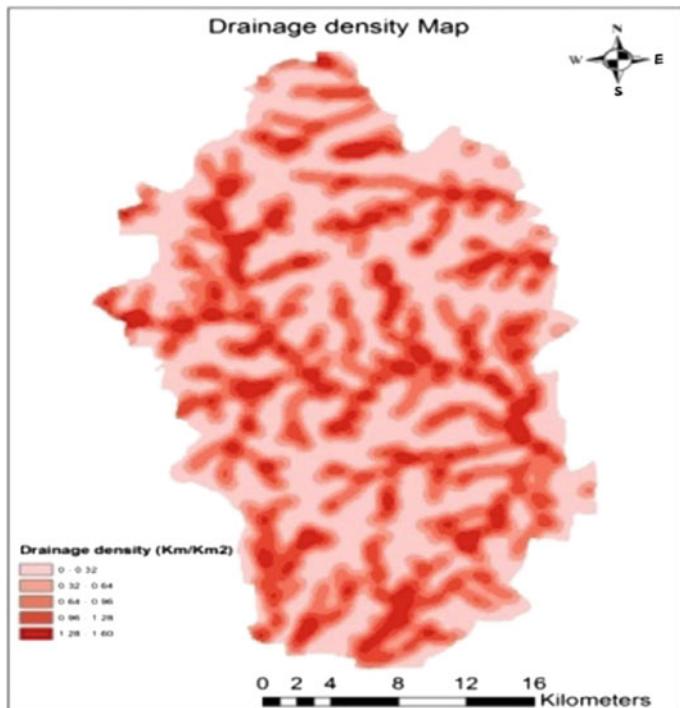


Fig. 4 Map for slope

Table 1 Reclassification

Parameter	Classes	Groundwater prospect (respectively with the classes in column 2)
Geomorphology	Denudational origin, anthropogenic terrain/origin, water body, structural origin	Very good, good, good, very poor
Slope classes	Nearly level (0° – 1°), very gently sloping (1° – 3°), gently sloping (3° – 5°), moderately sloping (5° – 10°), strong sloping (10° – 40°)	Very good, good, moderate, poor, very poor
Drainage density (km/km^2)	0–0.24, 0.24–0.48, 0.48–0.72, 0.72–0.96, 0.96–1.20	Very good, good, moderate, poor, very poor
Lineament density (km/km^2)	0–0.32, 0.32–0.64, 0.64–0.96, 0.96–1.28, 1.28–1.60	Very good, good, moderate, poor, very poor
Land use/land cover	Cropland, water body, river, scrubland, scrub forest, build land	Very good, good, good, poor, poor, very poor
Soil	Clay, red soil	Poor, very good

**Fig. 5** Map for drainage density

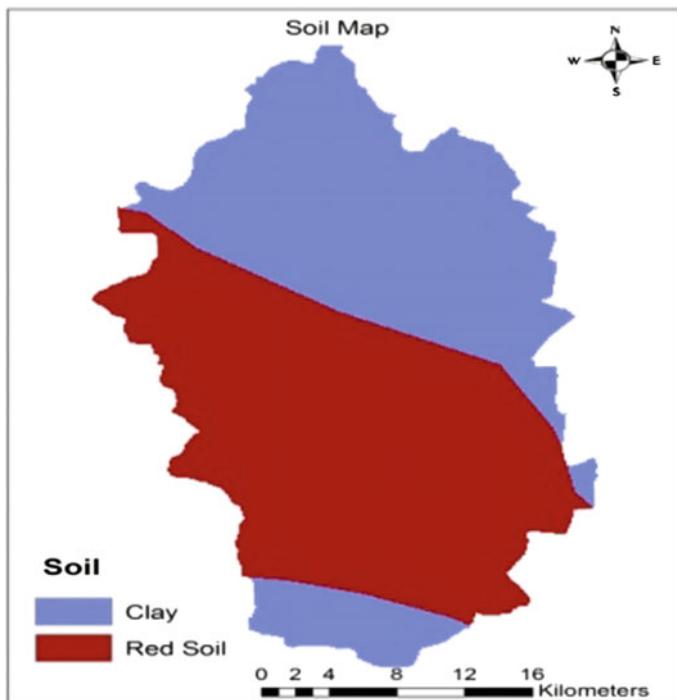


Fig. 6 Map for soil

VI. **Soil map:** It is shown in Fig. 6, and it shows that 45% of the study is covered by clay soil, while remaining 55% of study area is covered by red soil.

3.3 Reclassification

After creating thematic layers, these layers are reclassified as they all are in different units, in order to perform further processing so that these layers need to be in same unit. Here, all the vector layers are converted to raster formats. The reclassified map is shown in Table 1.

3.4 Normalization of the Weights Using AHP

The purpose of utilizing an analytic hierarchy process (AHP) is to recognize the favored option and furthermore decide the ranking of the options when all the criteria are taken into consideration [9]. The normalization steps executed are

Table 2 Comparison of class

Aspects	GM	DD	Slope	Soil	LD	LULC
GM	1	2	3	3	4	5
DD	0.5	1	2	3	4	4
Slope	0.33	0.5	1	2	3	2
Soil	0.33	0.33	0.5	1	2	3
LD	0.25	0.25	0.33	0.5	1	2
LULC	0.2	0.25	0.5	0.33	0.5	1
Σ of column matrix	2.61	4.33	7.33	9.83	14.5	17

Table 3 Normalized score table

Aspect	GM	DD	Slope	Soil	LD	LULC	\sum	Eigenvector	%
GM	0.38	0.46	0.41	0.31	0.28	0.29	2.13	0.35	35.5
DD	0.19	0.23	0.27	0.31	0.28	0.24	1.51	0.25	25.2
Slope	0.13	0.12	0.14	0.20	0.21	0.12	0.91	0.15	15.1
Soil	0.13	0.08	0.07	0.10	0.14	0.18	0.69	0.11	11.5
LD	0.10	0.06	0.05	0.05	0.07	0.12	0.44	0.07	7.2
LULC	0.08	0.06	0.07	0.03	0.03	0.06	0.33	0.05	5.5
Total	1.00	1.00	1.00	1.00	1.00	1.00	6.00	1.00	100.0

- I. The issue ought to be clearly characterized and afterward decayed into different topical layers having the distinctive classes of the particular subjects so that frame a system of the model.
- II. The relative significance qualities are resolved with Saaty's 1–9 scale [9] (Table 1), where 1 represents to equivalent significance among the two subjects. From the right side, the worth increment from 1 to 9 shows more significance, and toward the left side, the quality abatement from 1 to 1/9 demonstrates less significance.
- III. Table 2 demonstrates a matrix for comparing the classes in order to gain the precedence. A pairwise comparison matrix achieved with Saaty's 9 point significance scale. The corner to corner components of the matrix are dependably 1, and it just wants to top off the upper triangular matrix. In light of the verdict esteem by contrasting 1 thematic layer and other upper triangle matrix is filled. If a_{ij} is the element of row i column j of the matrix, then the lower diagonal is filled using this formula: $a_{ij} = 1/a_{ji}$.
- IV. Steps I to III as above gave us all variables needed to calculate priority vector. This is in notation done by finding out eigenvector in standardized form. We stick to the procedure for the same as illustrated in [10]. The results are as below (Tables 3 and 4).
- V. Principal eigenvalue, λ_{\max} , is calculated in Table 5. Saaty gives Consistency Index (CI) as degree of consistency using the following formula: $CI = (\lambda_{\max} - n) / (n - 1)$

Table 4 Ratio index of Saaty for values of n

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.89	1.12	1.24	1.32	1.41	1.45	1.49

Table 5 Calculation of principal eigenvalue

	Thematic map	Σ of column matrix (A)	Eigenvector (B)	$(A) \times (B)$
Geomorphology	2.61	0.35		0.9135
Drainage density	4.33	0.25		1.0825
Slope	7.33	0.15		1.0995
Soil	9.83	0.11		1.1796
Lineament density	14.5	0.07		1.015
LULC	17	0.05		1.02
Principal eigenvalue, λ_{\max}				6.3101

Table 6 AWR potential table of study area

Groundwater potentiality	Area (km/km ²)	Percentage (%)
Very poor	0.09	0.01
Poor	22.72	2.45
Moderated	185.05	19.93
Good	665.24	71.64
Very good	55.50	5.98

$n)/(n - 1)$, where λ_{\max} is the principal eigenvalue of the pairwise comparison matrix, and n is the number of classes.

- VI. Next step deals with finding consistency ratio (CR) which is a token of steadiness of pairwise comparison matrix. As per [11], it is calculated as, CR = Consistency Index/Ratio Index. Hence, CI = $(6.31 - 6)/(6 - 1) = 0.062$, that is CR = CI/RI = $0.062/1.24 = 0.05$ or $5 < 10\%$. Therefore, it is suitable.

3.5 Weighted Overlay Analysis

After the weightage of every major factor has been determined (Table 2), the weightage for the subclass of major factor has been allocated as given in Table 6. “Weighted Overlay” tool is subcomponent of Spatial Analyst Tools in ArcGIS is used for overlay analysis. Further, extraordinary units of each topic were allotted learning-based chain of importance of positioning from 1 to 5. Based on their hugeness with indication to groundwater prospects, where 1 represents poor prospects and 5 indicates phenomenal prospect of groundwater.

4 Observations and Discussions

The final weightages and rank of thematic layers and its individual features are shown as below. This consideration is based on Table 1 where we did reclassification. For every parameter, we considered the reclassified classes and labeled groundwater prospects like very good, good, moderate, poor and very poor, and then accordingly gave rank 5, 4, 3 ..., and so on, respectively.

1. For geomorphology, we got weights = 35.50%
2. For drainage density, we got weights = 25.20%
3. For slope, we got weights = 15.10%
4. For soil, we got weights = 11.50%
5. For lineament density, we got weights = 7.20%.

The AHP method provides an important aspect to justify the subjective evaluation of the problem. It is seen that the underlined study area has relatively large-area dimension suitable for the artificial groundwater recharge. The artificial groundwater recharge regions were achieved by combining selected six thematic layers as shown in Fig. 7. The separation of artificial groundwater recharge region was completed by weighted overlay technique in ArcGIS and finally assigned different recharge zones. The AHP is used to calculate the standardized weights of the separate themes and their dissimilar features. It is shown in Fig. 7 that few sections of center bit area are not suitable for artificial groundwater recharge site because of essence of auxiliary slopes, developed terrains with poor invasion limit. The groundwater recharge region map as shown in Fig. 7 pointed to that the excellent groundwater recharge region is concentrated in remaining parts of middle and western region because of the circulation of surge fields, alluvial fields, and drainage density with high penetration limit. This shows the geomorphology, slope, and waste thickness cover which assume an imperative part of groundwater development. Besides, the centralization of LULC and soil additionally encourages the invasion ability to the groundwater framework. About 0.01% area is under “very poor” region, 2.45% are under “poor” groundwater recharge region, 20% are under “moderated” region, 71.64% are under “good” region, and 6% are under the “very good” region. To conclude, the cumulative result of the weighted multi-influencing aspects through overlay analysis in GIS exposed the mapping of groundwater recharge region. Thus, it is evidently indicative that the outcomes are reliable with the results depend upon six factors, which control the groundwater recharge.

5 Conclusions

This study is an application of GIS and RS for multi-criteria analysis in resource evaluation and hydro-geo-morphological mapping for water resource management for selected *Kalmnoori taluka of Hingoli district (Maharashtra, India)*. The AHP

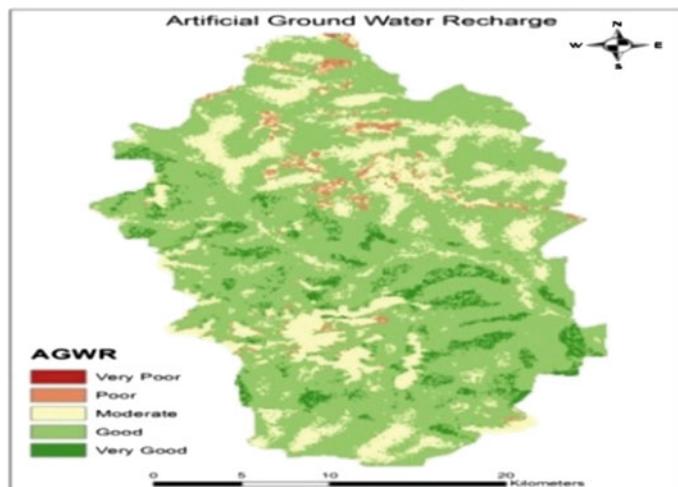


Fig. 7 Result map

process via Saaty's multi-criteria was evaluated for the variable values. The study has made it evident that the maximum of the study area can be a good candidate for artificial recharge sites. This study could be a role model for other similar geographic areas.

References

1. Rais, S., Javed, A.: Identification of artificial recharge sites in Manchi Basin, Eastern Rajasthan (India) using remote sensing and GIS techniques. *J. Geogr. Inf. Syst.* **6**, 162–175 (2014). <https://doi.org/10.4236/jgis.2014.62017>
2. Shankar, M.N.R., Mohan, G.: A GIS based hydrogeomorphic approach for identification of site-specific artificial-recharge techniques in the Deccan Volcanic Province. *J. Earth Syst. Sci.* **114**(5), 505–514 (2005). ISSN 0253-4126
3. Suhag, R.: Overview of ground water in India, Feb 2016. www.prssindia.org
4. Ministry of water resources government of India. Ground water scenario in India central ground water board, Jan 2016. www.cgwb.gov.in
5. Husen, S., Khamitkar, S.D., Deshmukh, N., et al.: Effective use of GIS based DSS for identification of artificial water recharge sites. In: ICI2TM-2018. ISBN 978-93-5254-640-5
6. Web resource at <http://www.bhuvan.nrsc.gov.in>
7. Government of India Ministry of Water Resources, Central Ground Water Board. Manual of Artificial Recharge Ground Water, Sept 2007
8. Web resource at <http://www.swat.tamu.edu>
9. Saaty, T.L.: Decision making with the analytic hierarchy process. *Int. J. Serv. Sci.* **1**(1), 83–98 (2008)
10. Web resource web.cjcu.edu.tw
11. Web resource www.ias.ac.in

Social Media Data Mining Techniques: A Survey



Dimple Tiwari and Manoj Kumar

Abstract Social Networking sites become very popular from last few decades. And that are become very useful for extracting the opinion of peoples regarding various things and topics. There are various techniques of data mining that are very much helpful for OSNs mining. In this survey, we focus on the concept of various algorithms and techniques of data mining that are used to mine online social network (OSNs), with special importance on latest topic of research area. There are several reasons which have made the study of OSNs gain importance by researchers. Here, we also look on the major issues that can occur at the time of mining social media data and how those are solved in last decades.

Keywords Online social networks (OSNs) · Data mining · Predictive technique · Descriptive technique · Classification · Clustering

1 Introduction

Due to the arrival of online social networks (OSNs), a change has occurred in the social interaction of people now-a-days [1]. Social media is defined as mobile based and the web based [2]. Social media sites produce a data that are extensive, turbulent, scattered, and active. Hence, data mining is a process of examine the immense, complicated, and dynamic OSNs data by various techniques. For understanding these techniques, first we have to understand what data mining is. It is a process that fetches and examines the data from large database, for discovering useful standard, shapes, and their correlations. It finds out the facts that are very useful. From the social media data, we can mine and analyze various things like student learning experiences,

D. Tiwari (✉) · M. Kumar

Ambedkar Institute of Advanced Communication Technologies and Research, Government of NCT of Delhi, New Delhi, India
e-mail: dimple.tiwari88@gmail.com

M. Kumar
e-mail: manojgaur@yahoo.com

opinion mining, toxicovigilance of medication abuse, user patterns, ranking social media users, culture-specific music learning behavior, positive customer reaction, topological data analysis, customer behavior, etc. This survey aim is to provide the information related to challenges occur in mining the social media data and how that are solved in last few years and provide look of various techniques and algorithms that become very useful in last year's to mine OSNs.

The rest of the survey is organized as follows: Section 2 provides challenges to mine social media networks (OSNs). Section 3 provides current status of OSNs research challenges. Section 4 provides data mining techniques. Section 5 provides predictive social media data mining techniques. Section 6 provides descriptive social media data mining techniques. Section 7 provides chronological summary of social media data mining techniques. Section 8 provides comparative analysis. Section 9 provides conclusion.

2 Challenges to Mine Online Social Networks (OSNs)

There are lots of problems occur when mining the online social networks (OSNs) [1]. These challenges have overcome in last decades by some mining techniques. But first look that challenges and then finds the techniques to solve them.

Opinion Mining: Mining the opinion of multiple users is very much difficult in social media research. Because most of the messages that are generated in micro blogging websites is textual information. Moreover, challenge with automatic opinion identification is that it is not remarkable to search two or more opinions in a one sentence [3]. **Group Detection:** Detecting communities in network has become a basic problem in network science [4]. Before finding the group anomaly, it is require finding single anomaly. **Recommender System:** Recommender system provides automatically suggestion to user that is according to their interest [5]. Collecting recommendation of large amount of OSNs data for items such as movies, gadgets, and books is a difficult process. **Link Prediction:** Web Application can be presented in graphs structure where nodes are interrelated. Exact prediction of new link and appearance of previous links are key problems [6]. So getting the link between them is an issue in research. **Finding an Expert:** There are various experts available on the OSNs related to particular domain. It is problematic who knows better relationships between OSNs make it challenging [7]. **Studying Trust and Distrust:** Now day's users are moved on to purchase on the basis of information that is available on OSNs. Decision of purchase is dependent on web, and there is developed enthusiasm to spread information but sometimes wrong information [8]. **Noise Removal:** Noisy data is meaningless data. It is very difficult and requires removing noise from data. In OSNs, data has come from different locations like from peoples, companies and group, etc., and it is very much unstructured. That not interpreted by machines correctly. So it is required to make data in structure. That can be received, read, stored, or changed in a proper manner (Fig. 1).

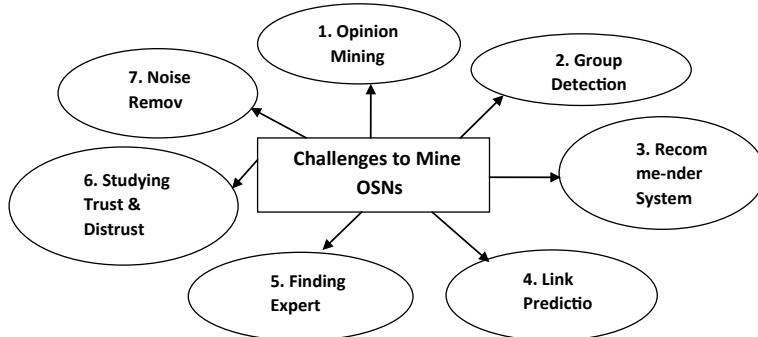


Fig. 1 Challenges to mine OSNs

3 Current Status of OSNs Research Challenges

3.1 *Opinions Mining*

Multi-model probabilistic graphical model is invented to solve the problem of multi-model aspect opinion mining in social media, apply on real-time data that collected from Flickr, Trip adviser and News articles [9]. The lexicon-based approach is used to sentiment analysis on the data that is collected from Twitter [10].

3.2 *Group Detection*

The biclustering method that is called BiP is used to calculate the related co-clusters. Its input can be giving either binary, discrete, or nominal values. It is one of the useful techniques for group detection [11]. GLAD method, stands for Group Latent Anomaly Detection, is a hierarchical Bayes model which is also suggested for detecting group anomalies [12].

3.3 *Recommender System*

It is to be suggested that social recommender with tags (cf-sn-tag) provides the best recommendation that is applied on dataset (Last.fm and MovieLens) [13]. Kaledio system is presented that trained by cluster-based machine learning mechanism, it learns automatically relationships among various context. And it achieves superior performance for recommender system [14].

3.4 Link Prediction

Covariance matrix adaptation evolution strategy (CMA-ES) is used to predict the links [15]. NeLP framework is invented, and experiments are showed that it predicts accurately positive with negative links on real-world social networks [16]. Structural signature-based approach is used for the link prediction problem, and one algorithm MTML Inter NeTwork Predictor (MINTP) is presented for predicting the links [17].

3.5 Finding Experts

A novel method matrix of graph regularized completion is presented for the issue of searching an expert, and it is applied on the data of popular question answering system, Quora. This approach improves the estimation for the missing values [18]. Topic candidate graphs are presented that are constructed by using only topic-relevant posts. It provides more statistical improvement and improves the performance of expert finding approaches. Dataset is collected from blogs [19].

3.6 Studying Trust and Distrust

New heuristic algorithm, H_OSTP, is applied on the Enron Email dataset, and it outperforms previous methods of trust path selection [20]. SWTrust framework is invented to create trusted graphs for evaluation of trust on OSNs. That implemented on data from Epinion.com and results showed high-quality trusted graphs [21].

3.7 Noise Removal

Brand tracking multi-faceted method is invented to collect representative data of huge OSNs content. A process of noise strain is moreover conducted [22].

4 Data Mining Techniques

Data mining techniques are divided into two categories [23] (Fig. 2).

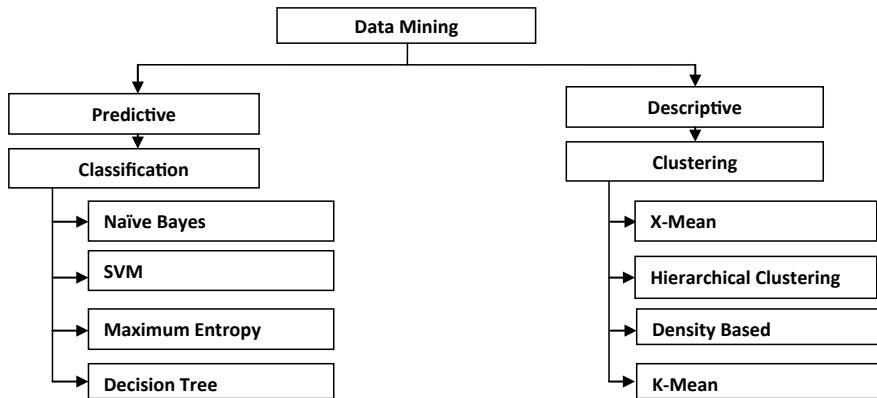


Fig. 2 Data mining techniques

5 Predictive Data Mining Techniques

Classification is a predictive technique. The main objective of this technique is to precisely forecast the target class. It has multiple algorithms that become very useful to mine OSNs data.

5.1 *Naïve Bayes*

It is an algorithm that is used in classification technique. Naïve Bayes classifier easily predicts the existence of a specific attribute in a class which is unrelated to the existence of any other attribute. Use purdue dataset and suggested that not extra human efforts are required in Naïve Bayes classifier it is very efficient in terms of computation time and training to a model not require any additional cost [24]. Collected data from Twitter API and classify the positive, negative and non-opinion tweets. When only Naïve Bayes technique is applied on that data so accuracy is 91% and after applying unigram Naïve Bayes, then it produces 90.17% accuracy [25].

5.2 *Support Vector Machine (SVM)*

This is supervised technique that is used for regression or classification difficulties. It used kernel trick to transmute data, and then according these transmutation, it fetches an ideal boundary between the outputs. Uses weight SVM technique with the help of Weka tool on Twitter data and provides accuracy of 81% [26]. Apply SVM on Twitter filtering and find out 87.5% accuracy [27].

5.3 Maximum Entropy

A maximum entropy classifier starts from making the assumption in terms of certainty. Used maximum entropy on Twitter data and achieve accuracy is 75% [26]. Collected census data from twitter crawling and suggested that maximum entropy technique is better than word count technique due to vocabulary that have enough English words that are negative and positive. This technique is proven to be effective in many natural language processing applications and sentiment classification of tweets [28].

5.4 Decision Tree

This is a tree structure that has a root node, leaf node, and branches. Collected data from Turkish facebook and apply Decision Tree multiple algorithms C5 accuracy is 47.54%, CART accuracy is 43.68%, CHAID accuracy is 41.40% [29]. Extract time-sensitive information from Twitter by online Twitter Monitoring System and apply various classification techniques after comparison best results are provided by the J48 Decision Tree method [30]. Collected privacy preference information from real social network users by facebook and apply three approaches (D-Tree-Active, Decision Tree, and Brute Force), and describe Decision Tree outperform Brute Force [31].

6 Descriptive Data Mining Techniques

Clustering is descriptive technique and also called unsupervised learning. It has various algorithms that are used in social media data mining.

6.1 X-means

Data is collected from the facebook with help of graph API and applied three clustering algorithm x-means, expectation maximization, and hierarchical clustering. Their result shows that x-mean and expectation maximization provides similar kinds of results with variation of many users [32].

6.2 *Hierarchical Clustering*

Various images crawled from instagram, and it is to be suggested that topological data analysis is more effective than k -mean and hierarchical clustering to mine high-dimensional social media data [33]. Suggested that hierarchical cluster not provides related results if data is not in hierarchical relation [32]. Del.icio from web pages and Last.fm from musical entries are two collaborative systems that are used. It describes that hierarchical agglomerative clustering is powerful to individualize navigational recommendation in collaborative tagging system [34].

6.3 *Density Based*

Dataset is collected from MySpace.com, then investigated density-based clustering algorithm, and created new scalable distance-based clustering technique. It is to be suggested that SDC outperforms the DBSCAN in opinion mining [35]. Collect dataset from Flickr and apply density clustering algorithm in geotag-related photographs, and achieve their analysis [36].

6.4 *K-mean*

K -mean algorithm on micro-blog data gathers from Twitter [37]. Use two clustering technique—one is simple k -mean and another is expectation–maximization and suggested these clustering methods are very useful in analyzing large data to extract characteristics of group that data is collected from online survey [23]. Collect picture data from Instagram crawls. After applying techniques, it is to be suggested that topological data examines outperforms k -mean for high-dimensional OSNs data mining [33].

7 Chronological Summary of Data Mining Techniques

To summarize the research, we have tried to list chronologically all the techniques of many authors along with algorithm and dataset used as shown in Table 1.

Table 1 Chronological summary

Year	Author	Technique	Algorithm	Other characteristics
2008	Andriy Shepitsen	Clustering	Hierarchical agglomerative clustering	Use Delicio from web pages and Last.fm dataset
2010	A. Selman Bozkir	Classification	Decision tree	Collected data from Turkish Facebook and apply Decision Tree multiple algorithms C5 accuracy is 47.54%, CART accuracy is 43.68%, CHAID accuracy is 41.40%
2010	Lujun Fang	Classification	Decision tree	Use dataset of Facebook and suggested Decision Tree outperforms Brute Force
2010	Tatsuya Fujisaka	Clustering	K-means	Micro-blog data gathers from Twitter
2011	Carlos Castillo	Classification	Decision tree	Extract time-sensitive information from Twitter and suggested best result achieve by the J48 Decision Tree method
2011	Christopher C	Clustering	Density based	Dataset is collected from MySpace.com and created new scalable distance based clustering technique. It is to be suggested that SDC outperforms the DBSCAN in opinion mining
2012	Deniele Quercia	Classification	Maximum entropy	Collected census data from twitter crawling and suggested that technique is better than word count technique. This technique is proven to be effective in many natural language processing applications and sentiment classification of tweets
2013	Xin Chen	Classification	Näive Bayes	Use purdue dataset and suggested that Näive Bayes classifier is very efficient in terms of computation time and training to a model

(continued)

Table 1 (continued)

Year	Author	Technique	Algorithm	Other characteristics
2013	Po-Wie Liang	Classification	Naïve Bayes	Collected data from Twitter API and apply Naïve Bayes technique on that data provide accuracy 91 %
2015	Jie Yin	Classification	SVM	Apply SVM on twitter filtering and find out 87.5% accuracy
2015	Memon	Clustering	Density based	Collect dataset from Flickr and apply density-based algorithm in geotag-related photographs and achieve their analysis
2016	Abeed Sarker	Classification	SVM	Accuracy is 81% using LibSVM and Weka tool
2016	Abeed Sarker	Classification	Maximum entropy accuracy	Accuracy is 75% using LibSVM and Weka tool
2017	Harish Kumar	Clustering	X-means	Data is collected from the facebook with help of graph API
2017	Harish Kumar	Clustering	Hierarchical cluster	Suggested hierarchical cluster not provide related results if data is not in hierarchical relation
2017	Veera Boonjing	Clustering	K-means	Use two clustering technique one is simple k -mean and second is expectation maximization and suggested these clustering methods are very useful in analyzing large data to extract characteristics of group that data is collected from online survey

8 Comparative Analysis

It is to be observed in this survey that classification technique means supervised learning is used more times than clustering unsupervised learning for mining the OSNs. Here, classification techniques like SVM, Naïve Bayes, maximum entropy provide proper accuracy in percentage. Instead of it clustering techniques like k -means, hierarchical clusters, hierarchical agglomerative clustering and density-based not provides accuracy in percentage format in this survey. It is to be proven that classifications maximum entropy technique is to be effective in many natural language processing applications and sentiment classification of tweets and Naïve Bayes is very efficient in computation time. Rather clustering k -means technique is very useful in analyzing large data to extract characteristics of group.

9 Conclusion

As Discussed, various techniques are available to mine the OSNs data. This survey describes predictive and descriptive techniques that both have various algorithms those can be used to mine OSNs data. APIs of the Facebook, Twitter, and many social websites provide easy data access to mine social media data. It is also discussed that various challenges can occur at the time of mining OSNs, but that can be solved by applying some methods. This survey paper shed light on various mining techniques that are proven very efficient to mine OSNs.

References

1. Nandi, G., Das, A.: A survey on using data mining techniques for online social network analysis. *Int. J. Comput. Sci. Issues (IJCSI)* **10**(6), 162 (2013)
2. Batrinca, B., Treleaven, P.C.: Social media analytics: a survey of techniques, tools and platforms. *Ai Soc.* **30**(1), 89–116 (2015)
3. Wilson, T., Wiebe, J., Hwa, R.: Recognizing strong and weak opinion clauses. *Comput. Intell.* **22**(2), 73–99 (2006)
4. Lancichinetti, A., Fortunato, S.: Community detection algorithms: a comparative analysis. *Phys. Rev. E* **80**(5), 056117 (2009)
5. Yang, X., Guo, Y., Liu, Y., Steck, H.: A survey of collaborative filtering based social recommender systems. *Comput. Commun.* **41**, 1–10 (2014)
6. Tylenda, T., Angelova, R., Bedathur, S.: Towards time-aware link prediction in evolving social networks. In: Proceedings of the 3rd Workshop on Social Network Mining and Analysis, p. 9. ACM, June 2009
7. Campbell, C.S., Maglio, P.P., Cozzi, A., Dom, B.: Expertise identification using email communications. In: Proceedings of the Twelfth International Conference on Information and Knowledge Management, pp. 528–531. ACM, Nov 2003
8. Guha, R., Kumar, R., Raghavan, P., Tomkins, A.: Propagation of trust and distrust. In: Proceedings of the 13th International Conference on World Wide Web, pp. 403–412. ACM, May 2004

9. Fang, Q., Xu, C., Sang, J., Hossain, M.S., Muhammad, G.: Word-of-mouth understanding: entity-centric multimodal aspect-opinion mining in social media. *IEEE Trans. Multimedia* **17**(12), 2281–2296 (2015)
10. Younis, E.M.: Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study. *Int. J. Comput. Appl.* **112**(5) (2015)
11. Balamane, A., Missaoui, R., Kwuida, L., Vaillancourt, J.: Descriptive group detection in two-mode data networks using biclustering. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 744–749. IEEE, Aug 2016
12. Yu, R., He, X., Liu, Y.: Glad: group anomaly detection in social media analysis. *ACM Trans. Knowl. Discov. Data (TKDD)* **10**(2), 18 (2015)
13. Zhou, J., Tang, M., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M., Lee, S.: Social network and tag sources based augmenting collaborative recommender system. *IEICE Trans. Inf. Syst.* **98**(4), 902–910 (2015)
14. Wu, C., Zhang, Y., Jia, J., Zhu, W.: Mobile contextual recommender system for online social media. *IEEE Trans. Mob. Comput.* **12**, 3403–3416 (2017)
15. Bliss, C.A., Frank, M.R., Danforth, C.M., Dodds, P.S.: An evolutionary algorithm approach to link prediction in dynamic social networks. *J. Comput. Sci.* **5**(5), 750–764 (2014)
16. Tang, J., Chang, S., Aggarwal, C., Liu, H.: Negative link prediction in social media. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 87–96. ACM, Feb 2015
17. Ahmad, M.A., Borbora, Z., Srivastava, J., Contractor, N.: Link prediction across multiple social networks. In: 2010 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 911–918. IEEE, Dec 2010
18. Zhao, Z., Zhang, L., He, X., Ng, W.: Expert finding for question answering via graph regularized matrix completion. *IEEE Trans. Knowl. Data Eng.* **27**(4), 993–1004 (2015)
19. Yeniterzi, R., Callan, J.: Constructing effective and efficient topic-specific authority networks for expert finding in social media. In: Proceedings of the First International Workshop on Social Media Retrieval and Analysis, pp. 45–50. ACM, July 2014
20. Liu, G., Wang, Y., Orgun, M. A., Lim, E.P.: A heuristic algorithm for trust-oriented service provider selection in complex social networks. In: 2010 IEEE International Conference on Services Computing (SCC), pp. 130–137. IEEE, July 2010
21. Jiang, W., Wang, G.: Swtrust: generating trusted graph for trust evaluation in online social networks. In: 2011 IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 320–327. IEEE, Nov 2011
22. Gao, Y., Wang, F., Luan, H., Chua, T.S.: Brand data gathering from live social media streams. In: Proceedings of International Conference on Multimedia Retrieval, p. 169. ACM, Apr 2014
23. Boonjing, V., Pimchangthong, D.: Data mining for positive customer reaction to advertising in social media. In: Information Technology for Management. Ongoing Research and Development, pp. 83–95. Springer, Cham (2017)
24. Chen, X., Vorvoreanu, M., Madhavan, K.: Mining social media data for understanding students' learning experiences. *IEEE Trans. Learn. Technol.* **7**(3), 246–259 (2014)
25. Liang, P.W., Dai, B.R.: Opinion mining on social media data. In: 2013 IEEE 14th International Conference on Mobile Data Management (MDM), vol. 2, pp. 91–96. IEEE, June 2013
26. Sarker, A., O'Connor, K., Ginn, R., Scotch, M., Smith, K., Malone, D., Gonzalez, G.: Social media mining for toxicovigilance: automatic monitoring of prescription medication abuse from Twitter. *Drug Saf.* **39**(3), 231–240 (2016)
27. Yin, J., Lampert, A., Cameron, M., Robinson, B., Power, R.: Using social media to enhance emergency situation awareness. *IEEE Intell. Syst.* **27**(6), 52–59 (2012)
28. Quercia, D., Ellis, J., Capra, L., Crowcroft, J.: Tracking gross community happiness from tweets. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, pp. 965–968. ACM, Feb 2012
29. Bozkir, A.S., Mazman, S.G., Sezer, E.A.: Identification of user patterns in social networks by data mining techniques: Facebook case. In: International Symposium on Information Management in a Changing World, pp. 145–153. Springer, Berlin, Heidelberg, Sept 2010

30. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 675–684. ACM, Mar 2011
31. Fang, L., LeFevre, K.: Privacy wizards for social networking sites. In: Proceedings of the 19th International Conference on World Wide Web, pp. 351–360. ACM, Apr 2010
32. Kumar, H., Kaur, H.: Clustering and ranking social media users based on temporal analysis. In: 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS), pp. 271–275. IEEE, Dec 2017
33. Almgren, K., Kim, M., Lee, J.: Mining social media data using topological data analysis. In: 2017 IEEE International Conference on Information Reuse and Integration (IRI), pp. 144–153. IEEE, Aug 2017
34. Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R.: Personalized recommendation in social tagging systems using hierarchical clustering. In: Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 259–266. ACM, Oct 2008
35. Yang, C.C., Ng, T.D.: Analyzing and visualizing web opinion development and social interactions with density-based clustering. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **41**(6), 1144–1155 (2011)
36. Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I., Chen, G.: Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Pers. Commun.* **80**(4), 1347–1362 (2015)
37. Fujisaka, T., Lee, R., Sumiya, K.: Discovery of user behavior patterns from geo-tagged microblogs. In: Proceedings of the 4th International Conference on Uniquitous Information Management and Communication, p. 36. ACM, Jan 2010

Automotive Vehicle-to-Everything (V2X) Communication Using IoT



P. Janani, Siddhant Verma, S. Natarajan and Aditya Kumar Sinha

Abstract In today's scenario, the automotive industry is upcoming with new technologies to compensate our current dependencies. In current scenario, it is seen to be undergoing series of transition as the industry is moving towards digitization and is therefore ahead to connected mobility. Therefore, the transports in today's generation have shifted from a conventional vehicle to intelligent vehicle which are equipped with advanced communication systems which alert or assist the driver. When the vehicles go online, it becomes the part of intelligent transportation system with advanced connectivity features. This upcoming technology is said to be creating a new era of connected vehicle services. Here in this paper, the implementation of V2V and V2I implementation of communication is using beacons (Bluetooth Low Energy) and LoRaWAN technology. This technology not only increases the safety aspects but also will provide ability to talk to nearby vehicles and infrastructures. Henceforth, this communication can enable drivers to act before in hand and respond to adverse conditions more precisely.

Keywords Vehicle-to-everything communication · Vehicle-to-vehicle communication · Vehicle-to-infrastructure communication · Bluetooth Low Energy protocol · LoRaWAN · Result analysis

P. Janani (✉) · S. Natarajan
Veltech University, Avadi, Tamil Nadu, India
e-mail: jananishri93@gmail.com

S. Natarajan
e-mail: drnatarajan@veltechuniv.edu.in

S. Verma · A. K. Sinha
CDAC, Pune, India
e-mail: Sid0308v@gmail.com

1 Introduction

With the upcoming technologies, the automotive industry is continuously bringing us new technologies. In the future, the automakers would surprise us with new technologies that will come into production. Wouldn't it be fascinating when if two neighbouring vehicles are talking to each other? But still nowadays, we do not have a reliable way for one car to communicate with another car to alert them in any exotic conditions. This is called V2X communication, i.e., vehicle-to-everything communication. More [1] precisely to define that vehicle will be able to talk automatically with nearby vehicles, architectures, pedestrians, etc. The communication that which we used till now was mostly cellular, so what about in future is still under process. The vehicles which are connected are considered to be a layer of network, and each sector of the layer is a domain in which it is connected to vehicle [2]. By this, it forms quiet an interesting way to see how two different ecosystems that evolved independently one on city side and one on the car side bring them together in an intelligent way that makes sense to the driver as well.

1.1 V2X Communication

Communication type insights:—The V2X communication piles up different types of communication which are widely used in various safety features such as autonomous cruise control, blind point detection and lane-changing assistance. Hence, this technology is said to dominate the market in future. Inclusion of V2I and V2V communication would be playing the major role in resolving critical situations such as congestion in traffic and will reduce emissions caused due to traffic. This [3] also paves new way that is convenient for e-parking and automated toll payments.

Connectivity type insights:—Nowadays, the V2X systems that are being installed in cars are mainly based on dedicated short-range connectivity (DSRC). The cellular connectivity has some drawbacks; BLE can be extensively used owing to its extensive coverage area; it requires considerable time for deployment. So, the upgraded hardware required to fulfil automotive industry requirements pertaining that V2X is expensive in nature.

Vehicle type insights:—The V2X communication technology is expected to have largest market as it takes the largest share in global automobile production in 2030. For example by the statistics of (OICA) Organization International Des Constructeur's automobiles, the CV sales are increased by 6% from 2015 to 2016; hence [4], the need is in high demand in the CV which would drive V2X market to high peak.

2 Problem Identification

This developing technology V2X communication is facing many hurdles to overcome nowadays. This is mainly due to some un-resolved question whether the minimum that is short and medium-range communication “bandwidth can be brought into existence to meet the needs of V2V and V2I [5]. Hence, preferably the main issue is said to be minimum range that would be possible to communicate within 150 m in any direction [6]. On the analysis of Government Accountability Office (GAO), most of the issues are listed as:

- Most of the information shared over the wireless medium of radio-frequency spectrum used by V2I communications is said to be not affecting the V2I performance.
- The mode of the feedback addressing states and local agencies is lagging of resources to maintain and deploy infrastructure technologies.
- Ensuring interoperability by developing technical standards.
- More looking forward to data security and privacy.
- To make sure that drivers respond appropriately to warnings.

3 Proposed Framework

The process is described as (Fig. 1),

- (1) For **Vehicle-To-Vehicle communication (V2V)**, beacons are used to advertise it and collect the needed information. (I am using Raspberry Pi to act as a beacon.) The beacons advertise themselves so frequently to its neighbouring vehicles and also will be able to scan the nearby vehicle/situations in its surrounding. For example, if vehicle 1 is travelling in gradual speed, then vehicle 2 is at the back on the same lane. In case if vehicle 2 approaches with high speed, it automatically indicates vehicle 1 that a vehicle is approaching at high speed, and in vehicle 2, an alert message is given that the vehicle is running at high speed. Hence [7], collision is avoided in this case, and both the vehicles pass without any interruption.
- (2) In the case of **vehicle-to-infrastructure (V2I)** communication [8], it can be done with the cloud through a gateway in bidirectional way. Here, LoRaWAN is used to connect with cloud. For example, (1) if the vehicle has to be parked in nearby area, it directly fetches information from the cloud through the gateway locating the available parking area nearby. For example, (2) the vehicle can directly talk with the approaching traffic signal and predict which lane is free to move further.

Bluetooth Low Energy

Bluetooth classic and Wi-Fi are the historic choices of people as it is most commonly

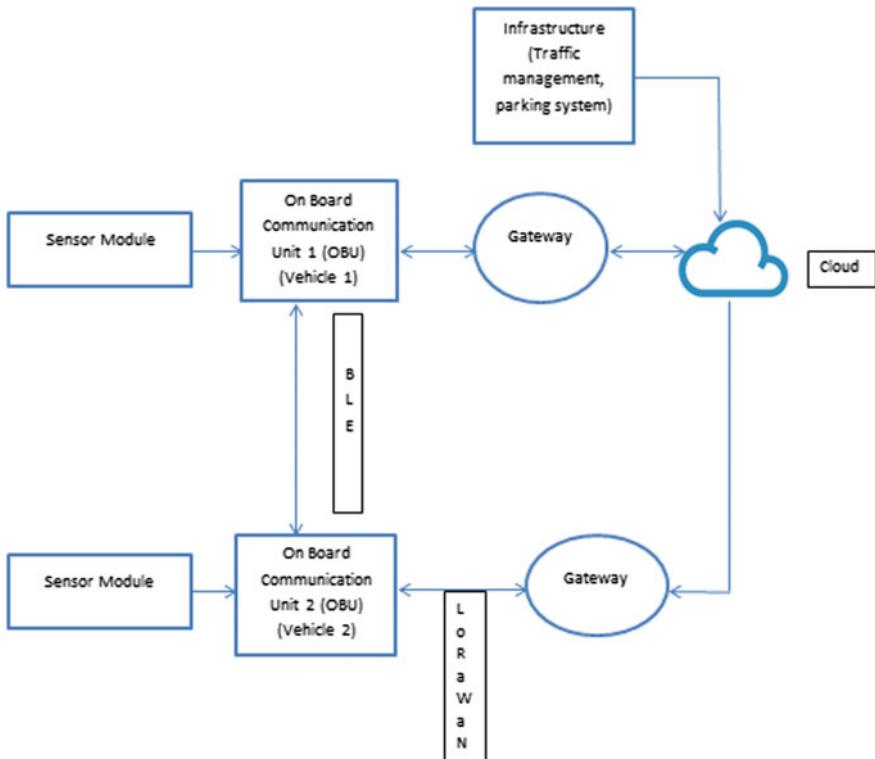
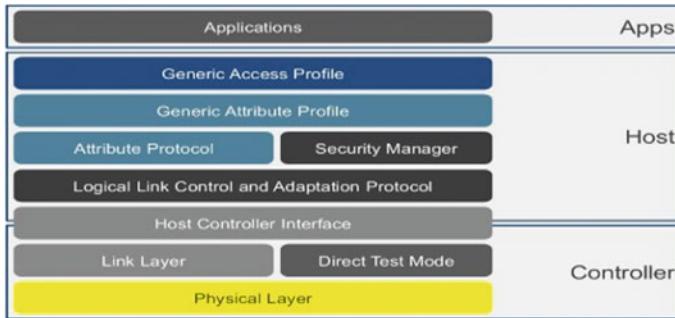
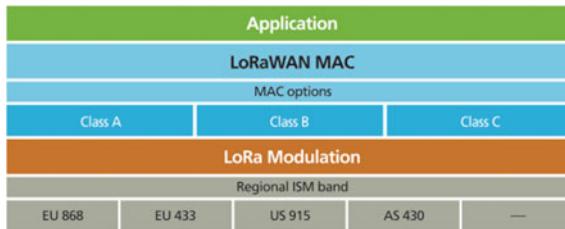


Fig. 1 Proposed framework

supported by all the smartphones. None of these connections are idealistic for IoT devices or applications. Normal Bluetooth classic although it supports large transfer of data, it is said that it consumes energy which in turn will easily drain small battery-powered devices which had to run for a long periods of time. Hence, BLE serves as the best that must operate for long periods of time on small energy sources, for example, coin cell batteries. The advantage in BLE is that its pairing is greatly simplified as it can advertise them easily and N number of devices can easily connect to the central server (Fig. 2).

LoRaWAN

LoRaWAN is defined as WAN [9]. Its features include at most very lower cost, in the case of IoT communication is said to be securable in bi-directional ways, (M2M) machine-to-machine, industrial applications and smart city. This favours to communicate over a distance of about 15 km. On the other hand, LoRaWAN consumes less battery life, bandwidth, features, etc. Key features of LoRaWAN include low power, geo-location, low cost, interoperability, long range (Fig. 3).

**Fig. 2** BLE protocol stack**Fig. 3** LoRaWAN architecture

4 Implementation of Proposed Framework

V2V Communication: Vehicle-to-Vehicle Communication. The beacon (Raspberry Pi) here advertises its details and also scans for the nearby devices. I am using a 32-byte beacon address such that first 8 bits consists of the manufacturer code, next 20 bits consists of the vehicle information, next 2 bits gives the type of vehicle, and last 2 bits gives the emergency situation. This communicates with the nearby vehicle through **Bluetooth Low Energy (BLE)** (Fig. 4).

UUID—This universally unique identifier (UUID) which is further split into five groups contains hyphens between them. It is a character set of 16 hexadecimal digits which is splitted into five sections of groups with a hyphen in between them. Its uniqueness is to mainly distinguish between iBeacons in network while comparing other beacons in network which are outside your control.

Major value—It is given by 16-bit unsigned integer. It is used to group the beacon that has the same proximity UUID.

Minor value—These values are designed to identify and distinguish from other beacons, given by 16-bit unsigned integer.

TX power—This is the broadcasting or transmits power with which a beacon is said to be broadcasting its signal.

E20A3911-444C-3443-4146-343934331111—16-byte beacon code

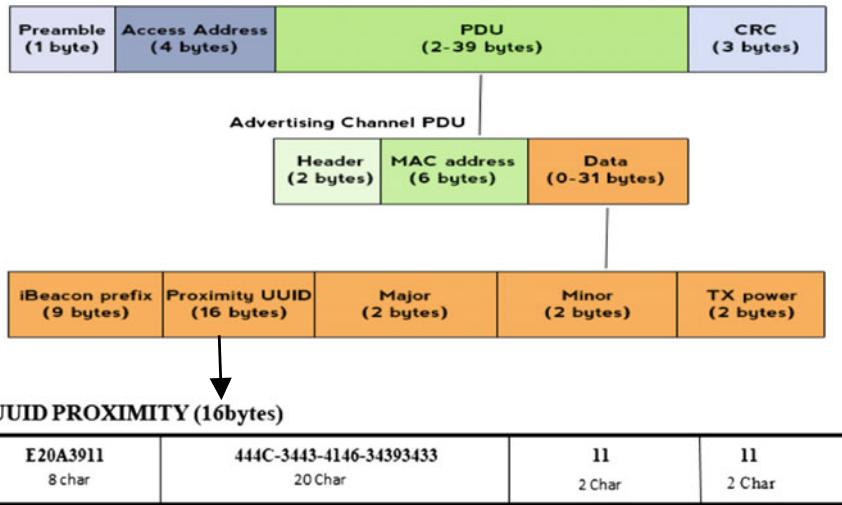


Fig. 4 Beacon data structure

(E20A3911)—First eight characters are manufacturer code in hexadecimal format.
 (444C-3443-4146-34393433)—Second 20 characters have vehicle no in hexadecimal format.

(11)—Last before two characters defines what type of vehicle is nearby like

10—CAR

11—BUS

12—TRUCK

13—EMERGENCY VEHICLE

14—TRAIN

15—MOBIKE

16—BICYCLE

(11)—Last two characters define the situation in the surrounding.

10—Normal Situation

11—Emergency Situation.

V2I Communication: Vehicle-to-Infrastructure Communication. Vehicles communicate with infrastructure such as traffic light signals, available parking lot nearby, etc., through **LoRaWAN**. It collects the desired information from the cloud through this protocol.

5 Result Analyses

V2V Communication

In V2V, the transmission of data is between the nearby or neighbouring vehicles. Early the manufacturers used dedicated short-range communication (DSRC). Its range was up to 300 m or 100 m or 10 s in highway speeds. Though it was established, it had some major problems like the radios are much expensive and is said that it is hard to obtain and the protocol is not finished after 15 years of work. Wherein the case of BLE (Bluetooth Low Energy) [10] its data rate is 10 kb/s, consumes less power, Frequency about 2.4 GHz. It is an open source and is of very ease in installation.

Scenario 1: To Avoid Collision

In case a line of cars in a highway are having high-speeds and imagine a situation on the road that one of the cars needs emergency breaks due to unforeseen event. By using the above V2V prototype using beacons through BLE protocol the message can be sent from that car to the neighbouring vehicle and that in turns passes on the message to other vehicles nearby. This continues until they will be able to take ease of action or perhaps emergency breaks in that lane. The main intention is to save lives and to take quick action precisely.

Scenario 2: Respond Quickly to Accidents

If suppose any vehicle had met with an accident outside the city where there is no mode of communication (hill stations), the beacon in that vehicle continuously advertises itself such that any vehicle passing that route would grab the data and further passes to other vehicles. This is done until it reaches the coverage area of RSUs [11] which then fetches the information and updates the central server. Hence, the needed help is sent to the provided accident zone.

Advantages

It provides the ability to see around corners.

It provides the ability to perform these actions in real time (Fig. 5).

V2I Communication

The V2I communication provides 360° of awareness of the similarly equipped vehicles. Hence, the driver remains in control with the vehicle as the driver will be alerted with any visual display or buzzer or a seat vibration technique.

Scenario 1—Public Transport Prioritization

My proposed work enables connected network V2I communication [12] to give the passengers real-time transit information so that they can predict when they can board the public transport or they have to wait next upcoming public transport. This enables to share robust data with each other. Such that passenger is informed before in hand about the next public transport which will be reaching at his/her location abruptly.

Scenario 2—Vehicle-to-Road-Side Units (RSUs)

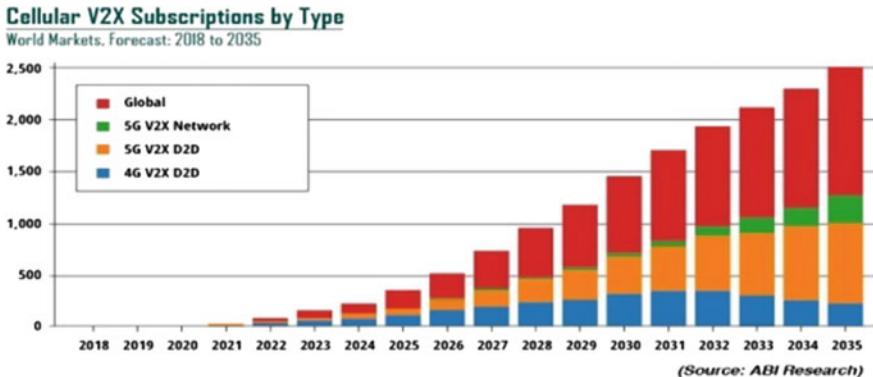


Fig. 5 V2X subscriptions by network

School Zone:—My proposal here in vehicle-to-infrastructure communication [13] finds a nearby school zone and intimates the driver to go slow in that lane. This information will be directed from the road-side units at necessary conditions.

Weather Conditions:—In case of bad weather conditions, road weather data from the vehicles are fetched and easily can be updated to traffic managing centres providing detailed real-time information to help, monitor and manage transportation. These in turn help in adjusting traffic lights and speed limits alerting the maintenance crews and notifying the motorists. Motorists before leaving their home will be notified by a pop-up message that would be most relevant to the current situation of the weather and traffic in their smartphones.

Scenario 3—City Information

In case if any construction work or road work is under process, it provides the warning to the driver accordingly like “WARNING: Roadwork ahead in 500 m” using my proposed architecture. In the cases of high classified building and marketing towers or shopping malls, it provides the new offers put up. Driver will be informed before in the case of intersection assistance [14], e.g., rail grid crossings. This will be like an alert or warning through any visual display or buzzer or alarm to the driver in situations when the driver is unable to see or hear the approaching vehicle.

Scenario 4—Parking Lot

In my architecture, V2I provides information hand to hand in your location from cloud in case of any vacant parking lots nearby, along with the best route to be taken to instead of going and checking that particular place. Adding to this in a more convenient manner, it provides the calculated distance from the current locus to the destination of the user. This also provides the precise route to be taken in order to reach the destination as soon as possible.

Scenario 5—Traffic Lights

Here from my architecture, the vehicles could receive the signals from intelligent road signs such as traffic lights. Intimates the driver to travel at 50 km/h to get

GREEN signal or It will become RED soon please slow down your vehicle or if it is YELLOW get ready to start. This technology grows day by day as it is said to influence the future market trends and automakers.

6 Conclusions

Thus, this proposed structure allows the progress of useful applications oriented to transportation systems [15]. This prototype mentioned above of V2V and V2I communications leads to adding benefits by reducing problems faced till now and minimizing the deficiencies that occurred while using these communication strategies separately. Hence,

1. Gather data related to the most visited zones. Reduce traffic jams and pollution.
2. The arrival time for emergency vehicles in the case of accident zones will be increased. Thereby, routing emergency vehicles to reach the zone faster by selecting routes has low density of vehicles.
3. The accuracy of data collected will be highly increased in case the RSU fail to collect the required data.

The above architecture proposal paves way for the betterment of traffic jam mitigation, as it quickly facilitates the calculation of new routes for smooth transportation modifying according to the data got in particular for heavy traffic conditions. All of this information can help authorities to decide future actions, improving the efficiency of transportation systems and reducing transportation costs. This is a key point since transportation issues highly affect the productivity and economic growth of developed countries.

References

1. Vehicle to Vehicle Interactions [Online]. Available at: http://www.mogi.bme.hu/TAMOP/jarmurendszerek_iranyitasa_angol/math-ch08.html
2. Bento, L.C., Parafita, R., Nunes, U.: Intelligent traffic management at intersections supported by V2V and V2I communications. In: 15th International IEEE Conference on Intelligent Transportation Systems. IEEE (2012)
3. Boskovich, S., Barth, M.: Vehicular network rerouting autonomy with a V2V, I2V and V2I communication matrix classification. In: 6th International IEEE Annual Conference on Intelligent Transportation. IEEE (2013)
4. Barrachina, J., Sanguesa, J.A., Fogue, M., Garrido, P., Martinez, F.J., Cano, J.-C., Calafate, C.T., Manzoni, P.: V2X-d: A Vehicular Destiny Estimation System that Combines V2V and V2I. IEEE (2013)
5. V2X Communication for Autonomous Vehicles [Online]. Available at: <https://www.digikey.com/en/articles/techzone/2017/apr/v2x-communication-for-autonomous-vehicles>
6. Transportation.gov U.S. Department of Transportation [Online]. Available at: <https://www.transportation.gov/automobiles>

7. Kim, B.W., Jung, S.Y.: Vehicle Positioning Scheme Using V2V and V2I Visible Light Communications. IEEE (2016)
8. Lai, W.K., Lin, M.-T., Lin, Y.-C., Jhuang, Y.-L.: An efficient packet forwarding mechanism based on bandwidth prediction with consideration of V2I and V2V communications. In: Sixth International Conference on Genetic and Evolutionary Computing. IEEE (2012)
9. Lauridsen, M., Vejlgaard, B., Kov'acs, I.Z., Nguyen, H., Mongensen, P.: Interference Measurements in the European 868 Mhz ISM Band with Focus on LoRa and SigFox. IEEE (2017)
10. Okamoto, E., Kunitomo, K., Akita, H., Kyo, T.: A Cooperative V2I Uplink Transmission Scheme Utilizing V2V Network. IEEE (2016)
11. Ni, Y., He, J., Cai, L., Bo, Y.: Delay Analysis and Message Delivery Strategy in Hybrid V2V/V2I Networks. IEEE (2016)
12. Future of Transportation [Online]. Available at: <https://www.youtube.com/watch?v=luo4my7yn58>
13. Vivek, N., Srikanth, S.V., Saurabh, P., Vamsi, T.P., Raju, K.: On field performance analysis of IEEE 802.11p and WAVE protocol stack for V2V & V2I communication. In: ICI-CES2014—S.A. Engineering College, Chennai, IEEE (2014)
14. Qualcomm-Future of Transportation [Online]. Available at: <https://www.youtube.com/watch?v=luo4my7yn58>
15. Securing V2X Communication [Online]. Available at: <https://www.youtube.com/watch?v=luo4my7yn58>

Summarization of Graph Using Question Answer Approach



Aditi P. Deshpande and C. N. Mahender

Abstract From past 400 years, charts, graphs, and other visual illustration of data have become a most important medium for communicating quantitative information (Giles et al. in Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 339–340, 2006 [1]). It contains important information and can provide essential knowledge that is not duplicated in other data formats. One important type of graph summarization is to produce small, and interactive question answer type summary based on user-based asked questions. Graphical representation, diagrams, charts, and graphs are main types of visual data in scientific domains. To detect, extract, and recognize text present in graphical are the important steps of the process. Generally textual data in charts/graphs includes axes titles, tick labels, legends, captions, notes. In general, OCR engines used, such as Tesseract OCR, ABBYY Fine Reader (Fei-Fei et al. in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 393–402, 2011 [2]), MODI (Microsoft Office Document Imaging) (Fomina and Vassilieva in Pattern Recogn Image Anal 23:139–144, 2013 [3]), are used to find and identify text in scanned document images. We used Tesseract OCR engine (Fei-Fei et al. in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 393–402, 2011 [2]) in our experiments. Particularly, the accuracy of recognition by Tesseract OCR is up to 97% when the scanned document images are used. Tesseract OCR also has some disadvantages like, it become failure in the detection of small text regions composed of one or a few words; such text regions are typical for chart images. In our proposed work, we try to resolve the failure outputs. After being cropped and deskewed, the detected text regions are passed to an OCR engine. When the proposed algorithm for text detection and localization is used as a preprocessing step, the experimental results show a significant increase in the recognition rate. For the work presented in this paper, the technical charts such as bar chart, pie chart are taken as input.

A. P. Deshpande (✉) · C. N. Mahender

Department of C.S.–I.T, Dr. BAMU, Aurangabad, Maharashtra, India

e-mail: aditi.deshpande96@gmail.com

C. N. Mahender

e-mail: nam.mah@gmail.com

Keywords Extraction of chart data · Question-answer generation system

1 Introduction

Summarization is the progression toward achieving the main ideas of the text in as less words as possible. It can be prepared in inscription, verbally, vocally, in words, by play, by composition performed in groups or individually. Wide range of research in the area of summarization shows that it is most popular and widely used strategies in science field. Summarization is a short and truthful illustration of given charts in such a way that the output contains the most important things of the main source in a reduced manner. Text summarization is a rising technique for perception, the main intention of any kind of documents.

Summarization is subfield of natural language processing. It has many applications like biomedical documents, weather news or news articles, education field, government offices, researcher's, electronic message summary, dumpy memorandum of news on portable device, information abstraction for businessman, and so on. There are many types of summarization like speech, audio, video, text summarization, etc. Automatic summarization is a way of lessening of a text document using readymade software, to form a summary with the most important points of the source document.

Text summarization techniques classified as extractive and abstractive text summarization [4]. Extractive summarization collects significant text from original data and groups them together without changing its meaning [4]. Abstractive summarization contains an understanding of source text with the help of linguistic technique to read the text and expressing it in own language. Most of the work has been done in foreign languages as well as Indian languages, like Hindi, Marathi, Panjabi, Sanskrit, etc.

Document-based summarization consists of single document or multi-document summary. Single document summary provides the suitable information involving in single document that helps the user in deciding whether the document is related to the topic of interest or not, and multi-document summarization creates synopsis of several source document. Generic or domain-specific summarization consists of all texts relevant to its topic or domain or condition. Generic summary does not target to any specific group; rather than it addresses spacious community of readers while Query. In the domain-specific text summarization, important task is to discover terms, phrases specific to the domain.

Query-based summarization technique consists of only answered the things which is asked by the user. These queries are generally natural language queries or keywords which are relevant to a particular area under discussion (Yogan and Ong 2016). Query-based document summarizer depends on correspondence of sentences and word frequency. Most of the work has been done on binary tree graphs and search trees and charts, etc. We are creating a summary of bar graph using automated question answering system.

We propose to use a data inference algorithm as a preprocessing step for text recognition with a common OCR system. From last 300 years, charts, graphs, and other visual data have been the important part in communicating the quantitative information. In proposed paper, we evaluate a process of data extraction from bar charts. We use automated extraction process in image processing. For the recognition of textual or arithmetical objects of every text section in the charts, a Tesseract OCR is used.

2 Related Work

The setback of perception of technical graphs has been addressed in an assortment of studies.

Chen [5] has explained a methodology to annotate every text strings in the graph with a semantic role, e.g., axis titles, legend, etc. without human intervention. For web-based diagram, search engine is described by Cafarella [6] for building an index, and the image of charts with its noted semantics is castoff [6]. The methodology does not endow with supplementary dispensation to take account of the unusual values of the chart as well, even though the annotator taking off the text strings and visual components for the task group [7]. As job completed by Elzer [8] and Giles [1], other methodologies of the values are dug out and improved, and on the other hand it solitary applied for black and white image or gray-scale image of charts [7]. They intend a specific structure [2] to examine and revamp charts is the study related to job is prepared by Savva [2]. They employ a technique to conjecture the original chart data ahead extracting graphics and text components [7]. The user is requisite to denote chart regions where text components are situated, where as the graphical components are extracted without human interference. The similar study explained that preset extraction and data revival procedure are described by Tan [9]. In mentioned system, they excerpt chart values depends on arrangement of textual and graphics components. The data extraction scheme of bar graph is explained by Huang in [10].

2.1 *Review of Related Work*

See Table 1.

2.2 *Comparison of Various Text Extraction Techniques*

Most of the researches have been worked on a range of text extraction techniques such as region based (CC based and edge based), texture based, morphological based,

Table 1 Review of research papers

Author and year	Paper title	Method	Rule	Limitations
Saracoglu and Alatan [12], Chaddha et al. (1994), Crandall [13], van der Schaar-Mitrean and de With (1998), LeBourgeois (1997)	Text segmentation in mixed-mode type images, systems and computers textual loom for indexing and repossession of image and video	Text detection method using texture features	Frequent color transition on image	Computationally complex when there are many edges
Sato et al. (1998)	Digital news collection needs video OCR, Proc. of IEEE practicum on content based admittance of image and video databases	Segmentation methods using character contours	Measure exact character boundaries	Sow may generate incomplete or distorted contour
Zhong et al. (1995), Lee and Kankanhalli (1995)	Text recognition in difficult color images, pattern recognition, zoning methods for handwritten character recognition	Character recognition methods using pixel connectivity	Use CC component	Fails to extract all the characters when there are joined or broken
Forssen and Lowe (2007), Matas et al. (2002)	Figure descriptors for MSER, vigorous large baseline stereo from MSER	MSER method	Remove non-text regions, threshold value needed	
Hase et al. (2001), Sakai et al. (1986)	Extraction of strings of characters from color documents, pattern recognition	Edge (text boundaries, strong edges), compression	Use kernel matrix, threshold value needed	Garbage values taken

Table 2 Comparison of various text extraction techniques

Techniques used	Images	Author and year	Remarks
CC method	Background is complex	Yao et al. (2007)	Each character pixel is supposed to have same color
Edge detection and k -means theorem	Images of signboard	Lai et al. (2008)	Efficient for illumination of odd
Discrete cosine transform and texture features extraction	Natural scene images	Angadi et al. (2010)	Ineffective when image background is more complex ex. trees, vehicles
CC based	Images of natural scene (devanagari text)	Raj et al. (2014)	Fails for small slanted/curved text
FFT domain filtering, SVM, k -means clustering	Images of textual scene	Azadboni et al. (2012)	Text characters having uniform color
Edge enhancement and CC method	Web-based images and caption text images	Zhang et al. (2012)	Not receptive to different types of background noises
CC based method	Bar chart images		Neighboring pixels which having change in color are labeled as dissimilar components

or mixture of these technique (i.e., hybrid approach). Researchers have used unusual type of images for carrying out tests. The comparative analysis of text extraction techniques is shown in Table 2.

3 Working of Proposed System

The proposed system contains three sections: visual components taking out section, text string extraction section, and the inference of data section [7].

The system is only build for 2D bar chart as source image and reproduces the data values as output. We have described the working of all modules below (Fig. 1).

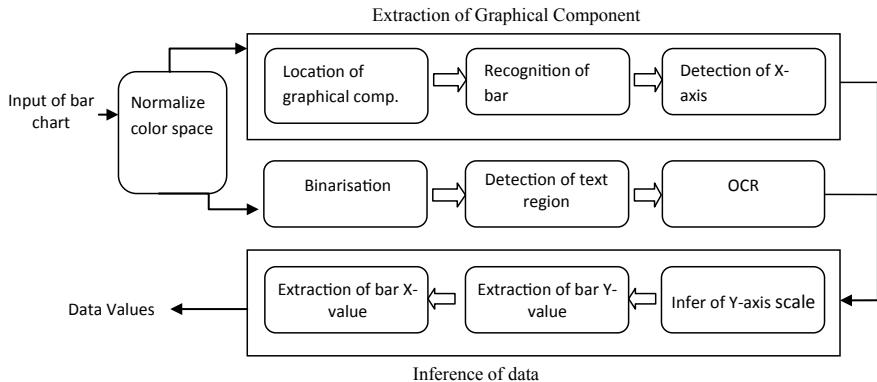


Fig. 1 Data extraction system overview [7]

3.1 Process of Data Extraction

3.1.1 Extracting Graphical Component

For perfect extraction of data, the bar extraction method follows some steps, first: identification of each bar distinctly, and second: identify as much bars as possible. We follow that the algorithm is alike to the one existing in [2] to recuperate the bars. We carry out connected component [7] tagging to the image. For the CC-routine and adjacent pixels which need to have alteration in shade are considered as dissimilar components [7]. When the CC routine is recovered, we spot the bar image constituents with the help of heuristics resulting from the graphical properties of the bars.

Every bar is defined by its place and height in pixels [7]. In turn to make out the x -axis, we assume that the x -axis is the horizontal line that all bars have on same edge [7]. A histogram of the base of the bar location is used to recover a familiar horizontal coordinate among the horizontal limits of altogether the bars of image; this is to be said as x -axis [1]. The background rectangles are painted white, while the factual bars are orange or red [1]. Background rectangles are usually colored as same color as neighboring rectangles, so they be close to other background rectangles [1]. On the other hand, bars are to be painted with dissimilar colors than adjoining rectangles, or gaps between bars [1]. If any one of the exterior points has the identical color as the inner of the rectangle, we classify the rectangle as piece of the background and remove it; otherwise, we organize it as a candidate bar [1].

3.1.2 Extracting Text Component

To recuperate the source values presented by the chart mechanism, the text strings coupled with the graphical modules from the chart image are significant rudiments

[7]. This scheme contains two fundamental steps [11]. The first step is to robotically trace text string areas from image. The subsequent phase is to implement OCR detection to text string components.

3.1.3 Text Regions Identification

For identifying text regions in chart [11], first is to mechanically trace out text regions and next is to apply OCR recognition. The image is already binarized. Removal of components is done whose region size is larger than the classic character size [7]. Then, to pinpoint which literacies epitomize only one expression, which is gone through isotropic dilation process per a tiny frame scope. It resolves and removes minute holes among pixel values that are nearest to each other only it should be adjacent [7]. Then, we are applying CC tagging to the dilated image. Through the tagging, it will be tagged the whole term as single component. Once the text region positions are given, the text region part conceded to the OCR step before dilation process on the images, it will preserve image superiority for the OCR. If the width of text areas is lesser than its altitude, it is perpendicular and mostly to the title of the y-axis.

3.1.4 Taking Out Textual String and Arithmetical Values

To recognize the numerical values of every text region, the Tesseract OCR is used. For any textual sections which produce blank spaces or only punctuation marks, the results like this are filtered out [7]. Finally, the texts which are recovered are represented as numerical values with their locations.

3.2 Preprocessing and Normalization

Preprocessing is needed to remove the redundant data and will focus on contrast of background.

Extraction of y-title is not possible in pass, and it needs second pass by rotating image in -90° . As we have done extraction process, we are getting some garbage values in the numerical data extraction like 00 or 1 g instead of 10. Whatever the unneeded garbage data is out, we directly delete it. And where the recognition problem is occurred because of the quality of image [7], sometimes it considered g instead of 0. At this stage, we have to assume it is 0 not g. The below standardizations are prepared concerning the charts:

- Charts should be 2D chart images.
- The color filled for the bars is a rock-solid color, to a certain extent of any pattern.
- The y-axis should follow a rectilinear ruler, reasonably than logarithmic scale.

- Alignment of y -axis is on the left side, and alignment of x -axis is on the foot scale of the chart image.

Some preprocessing is done with the chart image, such as color normalization and noise filtering. Noise-eliminating filters are used to the image for text identification in graphics [11]. Noise deduction filters such as bilateral filter. To transitory of chart to the extraction section, only preprocessing stage is changed the color space as a priority. The chart should be converted into grayscale version which is needed for the text extraction module.

Algorithm I A stepwise Algorithm for extracting numerical values and axes names and text strings from chart images [7].

Input Parameters: x-axis, y-axis, C, P
Output Parameters: Data values (data: titles; data: value), Y title, X title

```

1> if Pi.y is underneath x-axis then,
2> X <---- Pi
3> end
4> allocate X elements to X_title and X_labels depends on horizontal alliance
5> if Pi.x is to left of y-axis then
6> Y <---- Pi
7> end
8> allot Y elements to Y_title and Y_labels depends on vertical arrangement for Pi Y labels do
9> Scalei = Pi+1.y - Pi.y /Pi+1.value - Pi.value
10> end
11> _ = median(Scale);
12> Set datai.value <---- Ci.height * _;
13> Set datai.name <---- X_labelsi; where:;
X_labelsi x is contiguous to Ci.x
return : set of data pairs (data.titles, data.value)

```

4 Experiments

We have done the experiments on a total 10 bar charts, and those are taken from the net images sets [2]. OCR tool requires two pass for extraction of horizontal data and vertical data, i.e., in the first pass OCR extracts all x -values, title of graph, legends if they are presented in horizontal form. And in the second pass it extracts the y -scale value, i.e., all remaining vertical scale values.

The method shows the result of bars almost 80% in the first pass of image charts; if the bars were identified and located hundred percent, then also the almost 20% of y -scaling of these charts was not evaluated properly because of the drawback of the OCR tool [7]. Because of the image resolution and cleanliness of the bars, the maximum accuracies for the bar charts gained from the higher resolution images. Similarly, the y -scale estimated acceptably for utmost chart images apart from at what time outcomes of OCR are inaccurate [7]. For some images, more enrichment of the text region is obligatory to get concluding values for the data from images (Fig. 2). Displays a taster participation chart shows regions we are to be extracted.

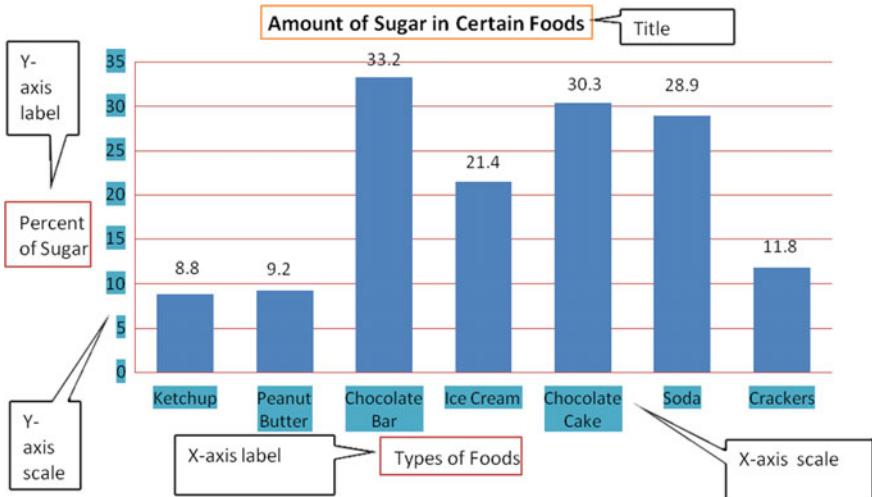


Fig. 2 Extraction of bar component

Additionally, some text improvement methods can be used to raise the precision of fallouts of the Tesseract OCR [7]. This should develop the exactness of y-scale withdrawal and the hatchets titles drawing out.

5 Question-Answer Generation System

Automatic question generation is subfield of natural language processing. Questions are used to reveal the informational needs. Graph-based methods are used for both single document and multi-document summarization. The answer is a request response or response to particular query. Answer is a thing that is done to deal with or as a response or reaction, statement, or situation. An answer to a particular question may express various form like, definition manner, list format, factoid based, descriptive manner, etc.

We have performed the five-point summary on the graph to extract particular values.

We have extracted the points of minimum and maximum values of x- and y-axis.

We have also extracted the minimum and maximum values of intersect points and then extracted the median point after that extract minimum quartile value that is Q1, and maximum quartile value Q3. In the quartile values, we calculate the 25% for Q1, 75% for Q3 (Table 3).

Finally, we put all above-calculated values in one matrix. Then ask query and answer will be the extracted values in the matrix.

Table 3 Question-answer

Questions	Answers
Q1. What data shows on x -axis?	– Height of trees in feet
Q2. How much gap between each data Intervals?	– 5 feet gap between each data intervals
Q3. What data shows on y -axis?	– Frequency distribution is given on y -axis
Q4. What is the title of graph?	– Height of black cherry trees
Q5. What is five-point summary of this graph?	<ul style="list-style-type: none"> – Min (x)-60, Max (x)-90, Min (Y)-0 – Max (Y)-10, median (x)-75, median (Y)-5, – Q1 (x)-65, Q3 (x)-85

6 Sample Question and Answer

See Fig. 3.

We have done experiment on above 10 numbers of images and will get answers as given in the chart. The method gives the results, on average 80% of bars in the first pass of image charts [7]. The y -scaling of about 20% of these charts was not evaluated properly because of the inaccuracies of the OCR tool [7]. Due to the image superiority and cleanliness of the bars, the highest accuracies is for the bar charts obtained from the high resolution images; though the bars were identified and situated correctly (Table 4).

Fig. 3 Sample input image
[14]

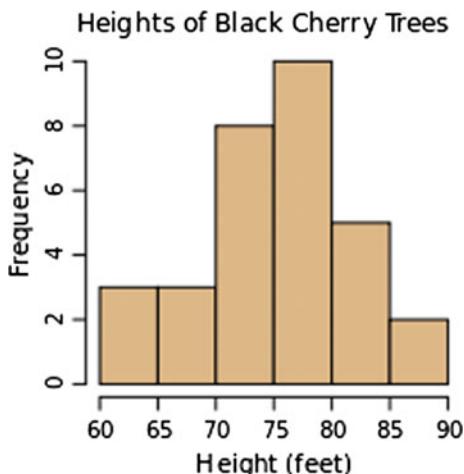


Table 4 Experimental results

Sample images	Q1. Min (x)	Q2. Min (Y)	Q3. Max (x)	Q4. Max (Y)	Q5. Median (x)	Q5. Median (Y)	Q6. Q1 (25%)	Q7. Q3 (75%)
Img1	0	0	–	70	–	35	15	55
Img2	0	0	2008	30,000	2004	15,000	6000	24,000
Img3	0	0	Others	200	Haryana	100	40	160
Img4	0	0	Crackers	35	Ice cream	17.5	7	28
Img5	0	0	Dec	100 k	July	50 k	20 k	80 k
Img6	0	0	–	0.1	–	0.05	0.03	0.07
Img7	0	0	Others	50	Calcium	25	6.5	43.7
Img8	0	0	Match5	70	Match3	35	15	55
Img9	0	0	2010	350	2005	175	43.7	306.5
Img10	0	0	7	20	3.2	10	2.5	17.5

7 Conclusion

In above paper, we used image processing methods to identify, read, and then extract data from 2D bar graph entrenched in digital papers or high resolution images. To improve that is valuable for semantic tagging of images, we give details of such a system for extraction of graphical and text string contents from bar charts to ask question and get answer as extracted values. An experimental outcome springs the 75–80% accuracy. The extracted values can be used in future studies for enlargement of specific knowledge sighting applications or improving query generation.

References

1. Giles, C.L., Liu, Y., Mitra, P., Bai, K.: Automatic extraction of table metadata from digital documents, In: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 339–340 (2006)
2. Fei-Fei, L., Agrawala, M., Savva, M., Kong, N., Chhajta, A., Heer, J.: Revision: automated classification, analysis and redesign of chart images. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. ACM, pp. 393–402 (2011)
3. Fomina, Y., Vassilieva, N.: Text detection in chart images. Pattern Recogn. Image Anal. **23**(1), 139–144 (2013)
4. Akter, S., Asa, A.S., Uddin, M.P., Hossain, M.D., Roy, S.K., Afjal, M.I.: An extractive text summarization technique for Bengali document(s) using K-means clustering algorithm. In: 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR) (2017)
5. Cafarella, M.J., Chen, S.Z., Adar, E.: Searching for statistical diagrams. In: Frontiers of Engineering, National Academy of Engineering, pp. 69–78 (2011)

6. Adar, E., Chen, Z., Cafarella, M.: Diagramflyer: a search engine for data-driven diagrams. In: Proceedings of the 24th International Conference on World Wide Web Companion, International World Wide Web Conferences Steering Committee, pp. 183–186 (2015)
7. Al-Zaidy, R.A., Giles, C.L.: Automatic extraction of data from bar charts. In: Proceedings of the 8th International Conference on Knowledge Capture. K-CAP, pp. 30:1–30:4 (2015)
8. Elzer, S., Chester, D.: Getting computers to see information graphics so users do not have to. In: Foundations of Intelligent Systems, pp. 660–668. Springer (2005)
9. Leow, W.K., Huang, W., Tan, C.L.: Associating text and graphics for scientific chart understanding. In: Proceedings of the Eighth International Conference on Document Analysis and Recognition, 2005, pp. 580–584 IEEE (2005)
10. Tan, C.L., Huang, W., Leow, W.K.: Model-based chart image recognition. In: Graphics Recognition, Recent Advances and Perspectives, pp. 87–99. Springer (2004)
11. Kasturi, R., Fletcher, L.A.: A robust algorithm for text string separation from mixed text/graphics images. *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(6), 910–918 (1988)
12. Saracoglu, A., Alatan, A.A.: Automatic video text localization and recognition. In: IEEE 14th Signal Processing and Communications Applications (2006)
13. Crandall, D.: Extraction of special effects caption text events from digital video. *Int. J. Doc. Anal. Recogn.* **5**(2–3), 138–157 (2003)
14. Atlas, by quartz. <http://atlas.qz.com/> (2015)

Analytical Assessment of Nature-Inspired Metaheuristic Algorithms to Elucidate Assembly Line Task Scheduling Problem



Shantanu Anandrao Lohi, Ajay A. Jaiswal, Harish V. Gorewar
and Snehal A. Lohi

Abstract Assembly line task scheduling (ALTS) or simple task scheduling (TS) is a process in which the cost of operations like time and resources is minimized by scheduling the required task amongst the units that are available at any given instance. In TS problem, problem formulation is carried out as a nonlinear constrained optimization and the facets of inequality and equality constraints are engaged to test the problem. In our paper, we consider the cuckoo search algorithm (CSA) and firefly algorithm (FA) for the implementation on a system of ALTS. By using the CS and FA, we try to obtain simulation results on the preliminary level, and then, they are compared to see their veracity of effectiveness on a group of tasks. Our paper focuses mainly on the preliminary investigations of usage of these two notable algorithms to gain an insight of their convention on scheduling activities effectively that are NP-hard normally. We also touch the issue of optimality required for exploration and exploitation in these metaheuristic algorithms. Our results indicate the efficacy of the cuckoo search algorithm and firefly algorithm and prove their potential to solve the ALTS problem in terms of multi-dimensional optimization problem.

Keywords Assembly line task scheduling · Metaheuristics · Firefly algorithm · Cuckoo search algorithm · Nature-inspired computing solutions

S. A. Lohi (✉)

Information Technology, Government College of Engineering, Amravati, India
e-mail: shantanulohi.kits@gmail.com

A. A. Jaiswal · H. V. Gorewar

Computer & Information Technology, KDK College of Engineering, Nagpur, India
e-mail: ajayjaiswal@yahoo.com

H. V. Gorewar

e-mail: harish.gorewar80@gmail.com

S. A. Lohi

Computer Science & Engineering, SRM College of Engineering for Women, Nagpur, India
e-mail: lohi.snehal@gmail.com

1 Introduction

One of the foremost issues faced by any resource allocation system is scheduling. This scheduling may be of resources needed to complete an intended output, sequence in usage of these resources, time required to complete a specific task, and so on. Scheduling is a universal phenomenon tackled by most of the systems in various fields irrespective of domains. More the advancements in the field of science and technology happen, more sophisticated our systems become, more efficient scheduling strategies are required to properly and effectively use the available resources at hand [1].

It is well-known fact that the optimization problems that are real world and real time in nature are quite difficult to solve. Furthermore, many such problems tend to be NP-hard as they progress with more input and are expected to solve more [1, 2]. Amongst new algorithms, cuckoo search and firefly algorithms are very notable because of higher efficiency they exhibit in solving complex problems. These types of problems require optimization without generating optimal results [3]. In assembly line operation part, a mathematics-based programming model based on multi-objectivity is considered in which one of the aims is to reduce or minimize any such functions specifying the usage of the resources used during specific time duration. This consideration is subjected to satisfying the limitation of resources and the technical limits of the operating constituents [4]. This is normally called as the economic dispatch for the operation of assembly line.

The assembly line task scheduling (ALTS) problem has complex and nonlinear features, and normally with the presence of inequality and equality constraints. With the advent of ALTS, many methods are incorporated to solve this problem, for example, interior points methods, linear and dynamic programming, iterative method, gradient-based methods, etc. With a view that a solution may get trapped in local minima, limitations crop up during optimization with regard to the use of conventional approaches [5, 6]. Incremental cost curves are required for the classical task scheduling algorithms, to be monotonically and linearly increasing, and some kind of estimation deems necessary to satisfy the requirements of the system. However, the delinquency still persists since the input/output individualities of modern assembly lines and generating units are highly nonlinear. Robot–human interfacing systems are also taken into consideration while addressing the issues of heuristics since robots may play bigger part in future [7].

When we considered the above-mentioned scenario, we were inclined to put our focus on using computational intelligence, nature-inspired computing methods [8], and metaheuristics for finding a feasible solution [9].

1.1 Motivation for Research

Assembly lines are indispensable activities in any production system. These are types of systems in which workstations are located in a serial mode and various tasks are

processed in a continuous manner. How can these assembly lines are designed and how efficiently they are implemented in order to minimize costing and enhancing the output still remains a crucial problem in various engineering and management fields [6].

Nature is an extremely inspiring entity that provides inspiration to develop computational algorithms also [10]. In today's world, most of the recently developed algorithms draw their inspiration from nature and natural phenomenon which in turn are closely related to humans in some or other way [11]. We, in a normal parlance, term them as nature-inspired and bears close resemblance to biological systems present in the nature [12]. Amongst these nature-inspired algorithms, we have a special kind of algorithms which are termed as swarm intelligence. Our reference algorithms, namely cuckoo search and firefly algorithms, are also the algorithms based on swarm intelligence [9, 13].

We are aware of other heuristic methods to be applied for this problem. These methods search for topographies of the problem and also look into other optimization methods available to us, thanks to the overwhelming research in nature-inspired computing [14, 15]. These special kinds of heuristics are known to us as metaheuristics. Metaheuristics are special because they are self-determining in their approach to reach to any feasible and acceptable solution for the problem at hand. For finding solution for optimization problems of unconstrained nature- and also population-based, various versions of metaheuristics are available now [16–18]. To accomplish a feasible yet efficient solution, population-based optimizing techniques are more effective [19]. In many applications, these are chosen since metaheuristic method allows the addition of constraints and moving forward towards solution [3, 5, 6].

By using population-based method, various new models of economic load balancing are deliberated upon. This problem can also be addressed by other methods like seeker optimization algorithm, particle swarm optimization, harmony search, ant colony optimization, genetic algorithm, water-cycle algorithm, fish school behaviour, social insect behaviour, and so on [18, 20, 21]. For obtaining the initial population by using a constructive method and also for improvement in the solution, we practised a local search technique [5, 22]. We try to obtain a solution of the considered population and use it in the information exchange within the groups of individuals. To achieve a solution that can be optimal, population is brought near to the process.

To solve the problem of TS, in this initial-level research, we tried our implementation with cuckoo search and firefly algorithms and performed some of the simpler comparisons. Our primary motivation in using these metaheuristic algorithms is the efficiency these algorithms demonstrate. Also, the results which are obtained by us are superior to other metaheuristics. A fine balance is accomplished for intensification and randomization [23]. We also receive a lower amount of control parameters.

Other reasons that are more obvious: a metaheuristic algorithm manages a nice balance in intense local search strategy and efficient exploration of the search space [11, 24]. This aspect normally makes the algorithm more efficient and enhances its ability to maintain a proper stability between exploitation and exploration of the resources and time of the system [8, 25].

The remaining sections of this work are organized as follows: the proposed techniques are explained in the following sections with simple implementation with simulation results and with the concluding thoughts of the paper.

2 Proposed Techniques

In this section, we are discussing our proposed techniques; here, only two techniques are proposed which are as follows.

2.1 Firefly Algorithm

We are using this algorithm for optimizing the assembly line task scheduling problem. Some of the ideas about the firefly algorithm are understood from Yang [1]. This algorithm is constructed on the behaviour of the fireflies. From the observed characteristics of this algorithm, we have three major rules with us, which are stated as follows:

- i. All the fireflies are considered to be unisex fireflies, and they are considered to get attracted to a brighter firefly regardless of sex of that bright firefly.
- ii. Value of the objective function is directly proportional to the brightness of the firefly.
- iii. Brighter firefly attracts less bright fireflies.

We understand that the less bright firefly will be attracted and move to the brighter firefly. The brightness can be affected while travelling to distances also. In case of no availability of a brighter firefly, random movement of the fireflies is observed in the space. This is the rule followed in the firefly algorithm.

2.2 Cuckoo Search Algorithm (CSA)

Cuckoo is a bird species. CSA is inspired by the strange, aggressive, and typical behaviour of cuckoos that they exhibit in their reproduction strategy and their special lifestyle. This algorithm was proposed by Yang and Deb [9]. Based on the brood parasitism, the CSA is an optimization algorithm based on the life of cuckoo species in which they lay their eggs in the already inhabited nests of other birds. Cuckoos usually remove eggs of the host bird in order to increase the probability of hatching of their own eggs.

The CSA contains a population of nests or eggs. This may be an objective function in this case. A single cuckoo egg in a nest/host nest represents a kind of initial solution and a would-be hatching/placed cuckoo egg represents a new solution. In case the

cuckoo egg bears a strong similarity with that of host bird's egg, then there is very less chance of cuckoo egg being discovered. Therefore, the fitness function can be linked to the alterations in the solutions [9].

The CSA rules are:

- The high-quality eggs in a best nest would be carrying over to the next inheritance cycle.
- One cuckoo is laying a single egg at a time and landfills it in a randomly but wisely chosen nest.
- P is the probability in which a foreign egg can be discovered by host. Host nest availability is fixed.

3 Implementation

In our paper, we are proposing two nature-inspired algorithms to satisfy the demand of saving the time and efficiently performing the task scheduling. The implementation is explained as follows.

3.1 Firefly Algorithm

The firefly algorithm consists of fireflies; in this case, a firefly is supposed to work on the system. The execution is similar to that of the algorithm here. In this case, firstly, the brightness of the fireflies is calculated which is considered to be a process of an allotted work [2]. Now, all the fireflies will search for the brighter fireflies. Once a brighter firefly is found, it attaches itself to it. In simple words, such works are combined and executed so that they do not collide and the speed of the execution increases along with usage of free resources. We consider here that the optimization is done. In this system, the firefly behaviour in the system is as follows:

Input to the technique: Tasks to be scheduled are N,

List of Mutually Independent Task T_{MI} ,

Time needed for each task as T_i ,

Max size of combined task (M)

The Output is as follows: Scheduled Task List (T_S)

Total Time needed for completion (T_C).

Procedure:

Step 1: Create a Dictionary/Index type of the list for each task in the system, which should be in the format like shown in the following figure where it contains the task number and the mutually exclusive tasks (Fig. 1).

Step 2: now, to form dynamic clusters of the task the following criteria is used.

Task in cluster K = All the task with mutually exclusive tasks.

Task number	Mutually exclusive task
-------------	-------------------------

Fig. 1 List of task in index format

In this step the clustering technique used is the Bisecting K-means

Step 3: Now select any of the random population from column 1, so that we get a set of nodes from column 2.

Step 4: in this step evaluation of the inner and the outer brightness is done, to evaluate the inner brightness the time needed for each of these task to run in parallel and is termed as (B_i). Whereas to evaluate outer brightness, the time needed for each of task in the cluster to evaluate which is termed as B_o .

now

If $B_i < B_o$

Pass B_i to output

Else

Pass B_o to output.

Step 5: Repeat 3 and 4 until K no. of iterations and select the output for the minimum solution.

Time for running = Max [time of each task in the list].

The above algorithm is used for achieving our objective; for example, let us consider a task list which is shown in Table 1.

Now to form the clusters, the system scans the mutually exclusive sets, and from the table, we can see that we have task 1 and 4 that can be in same cluster as their tasks match with each other. Now, their tightness is calculated, and it fits the criteria the clustering is successful. This is done for all the tasks in the system, and lastly, the output is given out. The output for the assembly line task scheduling is shown in Fig. 2.

As shown in Fig. 2, we can see the implemented firefly algorithm. As observed in the figure, we can see that we have few of the processes merged together to become

Table 1 Table showing the list of tasks to be executed

Task number	Mutually exclusive tasks
1	Scanning size, assigning resources
2	Scanning type
3	Rotating motor
4	Assigning resources

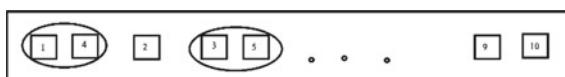


Fig. 2 Firefly-implemented belt

Fig. 3 Format of list of task

Task Number	Mutually Exclusive Task
-------------	-------------------------

one. Here, assume that we have 10 processes to execute. The importance of each process is calculated; the one with higher value are termed as the brighter fireflies. The other smaller fireflies will seek or will be attracted towards the brighter flies.

In Fig. 2, we can say that node 1 and 3 are brighter and are attracting the flies, but here only those flies are attracted which are opposite to each other, so the pairs formed are [1, 2, 8, 9, 14]; the distance or the difference between 1 2 3 is also sufficient for not merging. So now in this case, the time required is automatically low as multiple processes are executed at the same time.

3.2 Cuckoo Search Algorithm

In our second method, in order to increase the efficiency, the algorithm is imitating the cuckoo bird lifestyle. When a cuckoo bird is in the reproduction stage, to lay its eggs, it finds suitable nest from the surrounding and lays its eggs in it. After laying its eggs, the bird gets rid of the eggs already present in the nest of the bird. In this way, the competition is ended by the bird. The same process is implemented here. In this system, we have to optimize the system to be fast and high performance. Here, we have multiple works; now, some of the important works are termed as the nests and the smaller important works are termed as the eggs. The eggs should be laid into the nest to perform optimization of the system. This can be explained as follows:

To apply the cuckoo search algorithm.

In this algorithm, the first two steps are similar to those of the firefly algorithm.

Inputs technique: No. of tasks to be scheduled N,

List of Mutually Independent Task T_{MI} ,

Time needed for each task as T_i ,

Max size of combined task (M)

The output is as follows: Scheduled Task List (T_S),

Total Time needed for completion (T_C).

Procedure:

Step 1: Create a Dictionary/Index type of the list for each task in the system, which should be in the format like shown in the following figure where it contains the task number and the mutually exclusive tasks (Fig. 3).

Step 2: now, to form dynamic clusters of the task the following criterion is used.

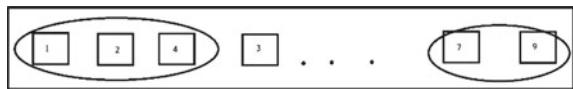
Task in cluster K = All tasks with mutually exclusive tasks.

In this step the clustering technique used is the Bisecting K-means,

Step 3: After initialization generate a set of “N” solutions, which are termed as the eggs in the nest.

Table 2 Result obtained

No. of tasks	Cuckoo search (in seconds)	Firefly algorithm (in seconds)	Improvement (in seconds)
10	39	39	0
20	150	154	2.9
50	217	220	1.3
100	2836	2748	3.2
500	3614	3566	1.3
800	6752	6822	1.02
1000	9746	9892	1.47
1500	37,889	37,840	0.12
2000	85,396	85,795	0.46

Fig. 4 Cuckoo-algorithm-implemented belt

Step 4: for each of the eggs in the set evaluate the total time needed together for solving to get the output of all tasks.

Step 5: now select a value “V” and discard all eggs whose fitness value is not in the top k results and replace them with the new results.

Step 6: Run for “i” iterations, after the iterations get the best solution from the list and discard the rest.

Step 7: Evaluate the time needed by the best solution.

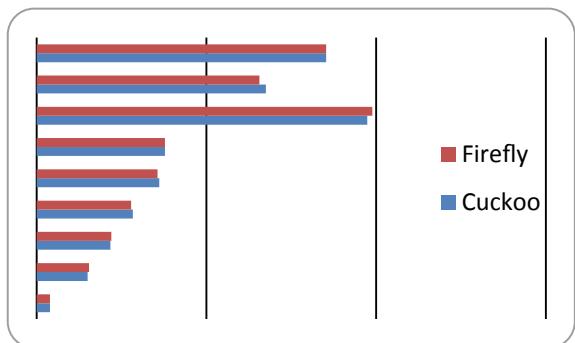
The above algorithm is used for achieving our aim; for example, let us consider a task list which is shown in Table 2.

As seen in Table 1, we have four tasks in general, since this is in the form of index we will start with directly with the generation of N solutions that is the eggs of the nest, once the solutions are found out the time evaluation is done for solving and the output for all the task. Now, a value V is selected which is a threshold; the eggs below this value are discarded, and new solutions are added. After I iterations, the best solution and the time are given.

As shown in Fig. 4, we can see the implemented cuckoo search algorithm. As observed in the figure, we can see that we have few processes merged together to become one. Here, assume that we have 10 processes to be executed as in the firefly algorithm. The importance of each process is calculated; the one with higher value is termed as the nest of the other species birds. The other smaller are termed as eggs and will seek or will be laid in the nests.

In Fig. 4, we can observe that there are two nests to be seen—one contains three eggs, and the other contains two eggs. The first nest consists of the {1, 2, 4} as the eggs; it is possible that 3 tries to be the part of nest and will remove any one of the eggs; similarly, we can see that in the nest two we have two eggs {7, 9} which can

Fig. 5 Comparison of time required by the proposed techniques



be added or removed with eggs. The efficient nests are only preserved in the system and are implemented further.

4 Result

In this paper, we have compared two of the most efficient techniques of the task scheduling which are firefly and cuckoo search techniques. The result is shown in Table 2.

In Table 2, we have applied both the techniques on a set of tasks. The tasks contain the time of execution also. For example, a set of task takes approximately 50 s when executed linearly but what happens when the algorithms are applied to it? As observed from the table, a set of 10 tasks can be optimized to 39 s by both the algorithms.

The calculation of the improvement is expressed by the following equation:

$$\text{Improvement} = 100 * (F - C)/F$$

where

F Time required by firefly algorithm

C Time required by cuckoo search algorithm.

From this, we can observe that when the number of tasks increases, the cuckoo search algorithm gives more efficient results. In the bar graph below, the time of execution is compared; as seen on the x -axis, we have the number of seconds required for the system to execute the process and the blue bar shows the cuckoo output and the red one is of the firefly. The y -axis shows the number of tasks to be executed (Fig. 5).

As shown in the figure, when the tasks are more, the time required by the cuckoo search algorithm is less than that of the firefly algorithm. For 2000 tasks, the time

Fig. 6 Performance improvement by applying the cuckoo search algorithm



required by cuckoo is nearly 400 s less which is equal to approximately 7 min which is a considerable difference for the output and the performance of the system. The improvement is shown in the graph below. The *x-axis shows the number of processes*, and the *y-axis shows the improvement of the technique over firefly* (Fig. 6).

5 Conclusion and Future Scope

This paper put forward the need for a technique that is important in helping the task of the assembly line task scheduling with minimum resources and less time with greater efficiency in performance. We establish from our study that the results obtained by the CSA are more viable in terms of resource friendliness and reduced time consumption than that of FA. This is obtained devoid of disturbing the performance of the system.

We wish to mention that the method we are proposing in this paper may not be efficient or effective than other methods being incorporated for solving these types of problems. However, one of the crucial aims of our research is to showcase how these two metaheuristic algorithms can be utilized in a proper and efficient manner in ALTS and how CSA outperforms FA in TS, which happens to be one the most studied and widely researched areas of engineering. Modifications are indeed necessary to give a diverse implementation strategy as per the requirements of ALTS, which may depend on the intended output and final product.

Futuristically, more investigations and research are necessary to establish the efficacy of the proposed mechanism in more complex and more automated environment of assembly lines and task scheduling where TS tends to be NP-hard after a certain limit. In our view, the performance of the both, CSA and FA, can be enhanced pertaining to ALTS or TS in a diverse environment of parallel tasks, simultaneous tasks, hybrid tasks, real-time task allocation, distributed TS environment, cloud-based approach of ALTS, and so on. We opine for more thorough implementation with diverse input and diverse environment. Further studies and analysis would be carried out, and the scope for further improvement is still there.

References

1. Yang, X.S.: *Nature-Inspired Metaheuristic Algorithms*. Luniver Press, London (2008)
2. Yang, X.S.: Firefly algorithms for multimodal optimization. Stochastic algorithms: foundations and applications. In: SAGA, Lecture Notes in Computer Science, pp. 169–178 (2009)
3. Fister, I., Jr., Yang, X.S., Fister, I., Brest, J., Fister, D.: A brief review of nature-inspired algorithms for optimization. [arXiv:1307.4186v1](https://arxiv.org/abs/1307.4186v1) (CS.NE) (2013)
4. Cui, H., Liu, X., Yu, T., Zhang, H., Fang, Y., Xia, Z.: Cloud service scheduling algorithm. *Res. Optim. Secur. Commun. Netw.* **2017**, 7 (2017). Article ID 2503153
5. Yang, D., Liu, Z., Zhou, J.: Chaos optimization algorithms based on chaotic maps with different probability distribution and search speed for global optimization. *Commun. Nonlinear Sci. Numer. Simul.* **19**(4), 1229–1246 (2014)
6. Dolgui, A., Gafarov, E.: Some new ideas for assembly line balancing research. In: IFAC (International Federation of Automatic Control) PapersOnLine **50**(1), 2255–2259 (2017). Elsevier
7. Chabria, S.A., Dharaskar, R.V.: Multimodal interface for disabled persons. *Int. J. Comput. Sci. Commun.*, 02–08 (2012)
8. Yang, X.S.: *Nature-Inspired Metaheuristic Algorithms*, 2nd edn. Luniver Press (2010)
9. Yang, X.S., Deb, S.: Cuckoo search via levy flights in nature & biologically inspired computing. In: World Congress on NaBIC, pp. 210–214. IEEE (2009)
10. Sur, C., Sharma, S., Shukla, A.: Egyptian vulture optimization algorithm—a new nature Inspired metaheuristics for knapsack problem. In: The 9th International Conference on Computing and Information Technology (IC2IT2013), pp. 227–237. Springer (2013)
11. Shi, Y.: An optimization algorithm based on Brainstorming process. *Int. J. Swarm Intell. Res. (IJSIR)* **2**(4), 35–62 (2011)
12. Ting, T., Man, K.L., Guan, S.U., Nayel, M., Wan, K.: Weightless swarm algorithm (WSA) for dynamic optimization problems. In: Network and Parallel Computing, pp. 508–515. Springer (2012)
13. Fister, I., Fister Jr., I., Yang, X.S., Brest, J.: A comprehensive review of firefly algorithms. *Swarm Evol. Comput.* **13**, 34–46 (2013)
14. Yang, X.S., Deb, S.: Eagle strategy using levy walk and firefly algorithms for stochastic optimization. In: *Nature Inspired Cooperative Strategies for Optimization (NICSO2010)*, pp. 101–111. Springer (2010)
15. Yang, X.S., Karamanoglu, M., He, X.: Multiobjective flower algorithm for optimization. *Proc. Comput. Sci.* **18**, 861–868 (2013)
16. Chen, H., Zhu, Y., Hu, K., He, X.: Hierarchical swarm model: a new approach to optimization. *Discr. Dyn. Nat. Soc.* (2010)
17. Civicioglu, P.: Artificial cooperative search algorithm for numerical optimization problems. *Inf. Sci.* **229**, 58–76 (2013)
18. de Paula Comellas Padro, F., Martinez Navarro, J., et al.: Bumblebees: a multiagent combinatorial optimization algorithm inspired by social insect behavior (2011)
19. Yang, X.S.: Firefly algorithm, stochastic test functions and design optimization. *Int. J. Bio-Inspired Comput.* **2**(2), 78–84 (2010)
20. de Lima Neto, F.B., Lins, A.J.C.C., Nascimento, A.I.S., Lima, M.P., et al.: A novel search algorithm based on fish school behavior. In: IEEE International Conference on Systems, Man and Cybernetics, (SMC), pp. 2646–2651 (2008)
21. Eskandar, H., Sadollah, A., Bahreininejad, A., Hamdi, M.: Water cycle algorithm—a novel metaheuristic optimization method for solving constrained engineering optimization problems. *Comput. Struct.* (2012)
22. Kaveh, A., Farhoudi, N.: A new optimization method: dolphin echolocation. *Adv. Eng. Softw.* **59**, 53–70 (2013)
23. Tang, R., Fong, S., Yang, X.S., Deb, S.: Wolf search algorithm with ephemeral memory. In: Seventh International Conference on Digital Information Management (ICDIM), pp. 165–172 (2012)

24. Tamura, K., Yasuda, K.: Spiral dynamics inspired optimization. *J. Adv. Comput. Intell. Intell. Inf.* **15**(8), 1116–1122 (2011)
25. Teodorovic, D., Dellorco, M.: Bee colony optimization—a cooperative learning approach to complex transportation problems. In: Advanced OR and AI Methods in Transportation, Proceedings of 16th Mini-EURO Conference and 10th Meeting of EWGT—Poznan: Publishing House of the Polish Operational and System Research, pp. 51–60 (2005)

#BiggBoss—Long-Run Event Detection and Sentiment Mining in Twitter



R. Geetha, P. Suthanthira Devi and S. Karthika

Abstract Online social media like Twitter and Facebook provides a common platform for presenting views and opinions of an individual on various events. This research work aims at detecting peak events that occur in online social networks using the proposed approach of exponential moving average algorithm and peak recognition method. A two-level hybrid sentiment analysis using n -grams and Naïve Bayes classifiers is performed on tweets to ensure the true sentiment of the user. The analysis was based on bag-of-words model and Bill McDonald's list for positive and negative words. The tweets were streamed for #BiggBoss and stored in sequential bins. The potential event was detected on the final day of result announcement by peak recognition method. The Naïve Bayes classifier predicted tweets with accuracy of 89% which would further aid in event summarization and eliminate event-related rumors.

Keywords Peak event detection · Hashtag · Temporal sentiment analysis · Twitter · #BiggBoss

1 Introduction

Twitter is the most happening social media platform among people who are interested to socialize digitally. This micro-blogging platform allows people from a society to post and share opinions on a specific topic with just 280 words. Since Twitter records real-time incidents, event detection and sentiment analysis are highly recommended

R. Geetha (✉) · P. Suthanthira Devi · S. Karthika
Department of Information Technology, Sri Sivasubramaniya Nadar College of Engineering,
Chennai 603110, Tamil Nadu, India
e-mail: geethkaajal@gmail.com

P. Suthanthira Devi
e-mail: devisaran2004@gmail.com

S. Karthika
e-mail: skarthika@ssn.edu.in

for identifying people's standpoint over a political party, crisis scenarios, entertainment, personal, and professional happenings. Identifying key events is a critical task because innumerable tweets get posted at a regular pace which is directly influenced by the duration of an event. Therefore, an approach of sequential binning is proposed to handle the complexity raised by the factor of time duration. Real-time tweets are streamed in successive bins depending on a time period assisting in identification of peak events for long-running events. The peak event description can be extracted by text summarization process which will automatically select the frequent terms from the sparse data.

The objective of this research work is to identify peak events and provide a contextual sentiment of the user for any long-running events. A hybrid sentiment analysis system is designed for predicting the true sentiment of the user about that particular event. A novel aggregation of n -grams and Naïve Bayes classifier is selected for predicting the sentiment subsequently identifies the sentiment associated with a tweet. Collective analysis of the sentiment predicted would further help in event summarization. The time lapse plays a key role in identifying the peak event because of the factor that as the news gets old, it loses its essentiality among audience. The results of peak event detection and sentiment analysis will be a key factor in decision making, text summarization, and user behavior analysis.

2 Related Work

Content Identification for Event. A specialized journalistic tool was introduced by Becker et al. [1] to benefit journalists to gather original information in real time without delay. Social sensing feature was found which aided the tweet identification for journalism [2]. The topics needed for journalists must be rich in content, sensitive and genuine. Information from all popular social media sites like Facebook and YouTube will be extracted by a tool to supply extensive news feeds for the journalists [3]. Truthfulness of the news collected is most important factor as it denotes the reputation of a journalist. The faster the reporting of news the much better appreciation and recognition a journalist receives [4].

Sentiment Analysis. Schifferes et al. [5] showcased the challenges in analyzing the sentiment of a message that arises due to the informality and multilingualism that exists in user's message. The complexities are handled by considering the emotion expressed from the words in the tweet such as sad, happy, angry, and many more. The approach proposed in [6] uses graph propagation algorithm and plots the emotion tokens extracted from tweets to different polarities. The sentiment analysis algorithm is applied to the generated graph to classify the emotion tokens. One major issue in the proposed approach is the difficulty in finding sentiment of tweets that are streamed for a long time. The emotion prediction performs well than lexicon analysis because the later one will be time-dependent [7].

Twitter Attributes Analysis. The authors of [8] proposed a novel approach for predicting the social context of user reviews by building hypothesis based on the

constraints like author consistency, trust consistency, co-citation consistency and link consistency. The retweet behavior of the users is a way that help people start communicating with each other depending on various reasons with different styles [9]. A novel concept for tracking large scale events was proposed by [10] with approaches that analyze tweet attributes for streaming topics and visualizing it over a range of time and user activity.

3 Methodology

The peak event detection and sentiment analysis algorithm are applied to the streamed, preprocessed tweets after tweet stream analysis process. The detailed architecture of the system is depicted in Fig. 1. The temporal peaks which can be a significant peak event are identified by finding out the tweet frequency. The tweets posted in each time period are extracted and stored in bins based on the tweet arrival rate. The measures like mean, mean deviation, and difference are calculated for analyzing the bin-wise stored tweets to find the peak in tweet timeline. The most frequent terms are collectively given a label when a peak is detected. This peak event is dynamically varying because of the probability of higher number of incoming tweets for each bin and the window is slid accordingly for peak detection process. The main objective of employing hybrid sentiment analysis is to arrive at the truest sentiment among all other predicted sentiment. The bag-of-words model representation is employed for identifying the most frequently occurring n -grams which help in predicting the user's intention and standpoint based on the streamed tweets which are later helpful in annotating the event. The process of event detection and hybrid sentiment analysis is briefed as follows.

Tweet Collection and Preprocessing. Tweets are collected from Twitter using Twitter API based on the string query and stored such that each bin has tweets of a particular day. The tweets posted by the users are generally unstructured which needs to be cleaned and formatted for context-wise and content-wise analysis. Therefore, a set of preprocessing rules are applied on tweets like removing hashtags, punctuations, URLs, numbers, stop words, and emoticons. Finally, all the remaining content is converted to lowercase, spell-checked, corrected, and stored in a repository. The preprocessed tweets with the username and time are considered for the next step of tweet stream analysis.

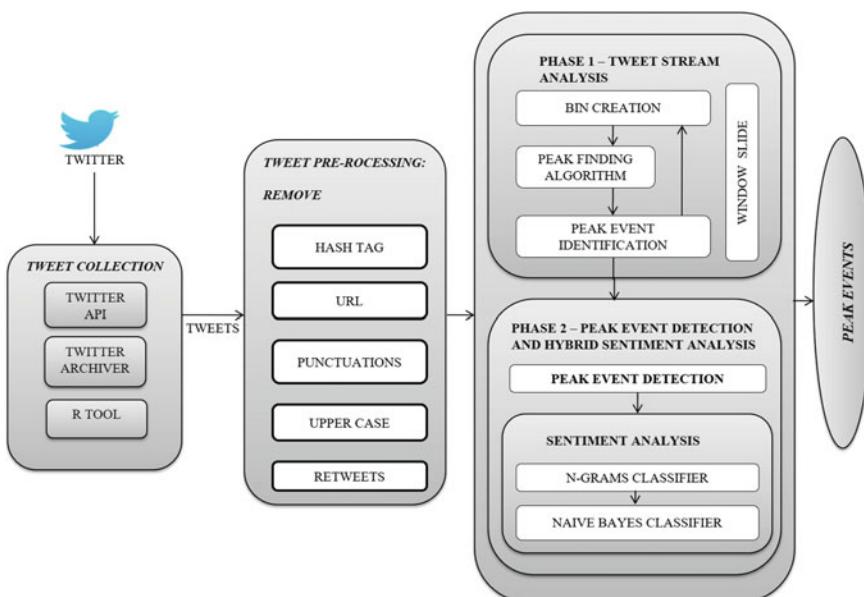
Tweet Stream Analysis. The preprocessed tweets along with the timeline parameters like date and time from the sequential bins are considered for computing mean and mean deviation. The mean and mean deviation are computed in the peak finding algorithm as shown below in Table 1.

The mean value for Day 1 is computed and is compared with the mean value of Day 2 to obtain mean deviation that has occurred in the time interval of Day 1 and Day 2. Initially, the peak is set at Day 2 and is dynamically assigned based on the following criteria. The new peak is detected if the product of the current bin's tweet count and τ (where $\tau = 2$) is higher than the current mean. Initially, the bin counts are

Table 1 Peak recognition algorithm

Step 1. Find peak_windows(C)
Step 2. Initialize first bin count mean as mean = C_1 and bin as windows = [].
Step 3. Compute mean deviation of bins as mean_dev = variance (C_1, \dots, C_p).
Step 4. If mean_deviation > bin count, then form new peak window for($i = 2$; $i < \text{len}(C)$; $i++$) if ($ C_i - \text{mean} / (\text{meandev}) > \tau$ and $(C_i > C_{i-1})$, start = $i-1$
Step 5. Update mean and mean deviation values. update(mean, meandev, C_i)
update(mean, meandev,c)
Step 1. diff= old mean – update value
Step 2. newmeandev= $\alpha * \text{diff} + (1-\alpha) * \text{oldmeandev}$ (where, α – constant)
Step 3. newmean= $\alpha * \text{updatevalue} + (1-\alpha) * \text{oldmean}$

considered for computing the tweet arrival rate for each time period. The historical weighted running average of tweet rate is decided if the tweet rates are higher than the mean tweet rate. This will recursively identify the local maxima of tweet rate and employ a window slide to track the peak tweet rate (Fig. 2).

**Fig. 1** System design of the proposed event detection and sentiment analysis system

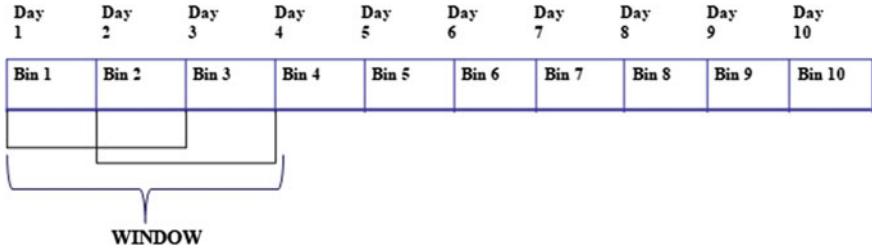


Fig. 2 Bins for each day and the window slides based on the peak detected at each iteration with the assumption of 1 bin per day

Event Detection. The event detection is performed by determining key features like start and finish window in the bin list and the unigram event's keywords. The event keywords are generated by constructing the term frequency-inverse document frequency (TF-IDF) for all tweets in the window. The labels for window are determined by ranking the event keyword's frequency obtained in TF-IDF ranking method. The peak events are detected and marked to be a significant event occurrence regarding the topic considered. This method of ranking and shifting the peak event depending on the peak recognition is defined as exponential moving average algorithm.

Hybrid Sentiment Analysis. The sentiment for event is detected by the proposed hybrid approach by classifying tweets in two levels. Initially, the preprocessed tweets are classified using n -grams and bag-of-words model, and the sentiment for tweets are obtained. The next level of sentiment classification takes the resultant neutral tweets from the n -gram classifier and applies Naïve Bayes classifier. This recursive classification approach is carried out to find the truest sentiment and overcome the misclassifications in the n -gram approach. The positive and negative word lists are taken from McDonald's list, and the sentiment score is computed for each tweet.

Rules for Classification. If the word is present in the positive words list, then the word is declared to be positive. Similarly, a word is determined to be a negative word by comparing the negative word list from McDonald's word list. If the word is neither present in positive nor negative list, it is declared to be a neutral word. The sentiment score is computed by the following expression and summing the score for each word in a tweet. Let $PW = \{\text{Set of Positive Words}\}$, $NW = \{\text{Set of Negative Words}\}$. Then, the list of words score SW is calculated as:

$$\text{Score}(W) = \begin{cases} 1, & (Wx \in W)(t \in T)(Wx \in PW) \\ -1, & (Wy \in W)(t \in T)(Wy \in NW) \\ 0, & (Wz \in W)(t \in T)(Wz \in PW)(Wz \in NW) \end{cases} \quad (1)$$

Table 2 Tweets collected from Twitter Archiver for query string ‘BiggBoss’

Date	Time	Username	Tweet_Text
9-26-2017	21:37:09	@Ashmithacutie Rani Ashmitha	@HarishTribe please...guys vote for Harish...! he s the only person capable to win the BIG BOSS title
9-29-2017	19:09:21	@Priyaa Sweetz Riya Karthik @vijaytelevision	If #Snehan Wins the Title, plz #dontStart Big Boss next season ever. We hate Snehan to the Core
9-29-2017	21:35:05	@AnushaRamesh2Anusha Ramesh	Seriously someone give him a trophy from big boss to snehan..let him stop crying #BiggBossTamil
9-30-2017	0:01:18	@Oviya_addictsOviya Army	If Oviya is still in Big Boss ! ! ! Not these 4 Mens; Even 40 Mens can't win her...?? RT ?? (If you agree)... https://t.co/IHwRekj7Gl

4 Results and Discussions

Dataset. Twitter is the most frequently used micro-blogging site with a limitation of just 280 characters which makes the text and context much crispier and comprehensible for the user to present and observe. The dataset for the research work was streamed for the query string ‘BiggBoss,’ which happened to be one of the top trending topics during late 2017. A total of 5290 tweets related to the keyword ‘BiggBoss’ was collected for 21 days specifically from September 16, 2017, to October 6, 2017. Table 2 represents the tweets streamed from Twitter API using Twitter Archiver for the keyword ‘BiggBoss’ and stored bin-wise with metadata like ‘date,’ ‘time,’ ‘username,’ and ‘tweet text.’

Tweet Stream Analysis. Table 2 presents the peak finding algorithm worked out for all bins. The mean values for the tweets in bins and the values of mean deviation found in the subsequent bins are calculated for twenty-one-day timeline.

Event Detection. The peak event is detected by recursively going through the peak finding algorithm for an event. From the algorithm, the peak event is determined based on the update value function. From the dataset taken for analysis, it is found that the peak event occurred on Day 16 which was influenced due to the difference (419 tweets) achieved by total tweets posted by users on that particular day. The highly critical events which are forming peaks and evolving from the time of ‘BiggBoss’ event occurrence are presented in Fig. 3 and Table 3.

Hybrid Sentiment Analysis. The bag-of-words model and n -gram approach are the initial steps in hybrid sentiment analysis proposed in this research work. The most frequently used words for the event ‘BiggBoss’ are filtered, and a word cloud is formed as shown in Fig. 4. The results of n -gram approach yield three different sentiments—positive, negative, and neutral.

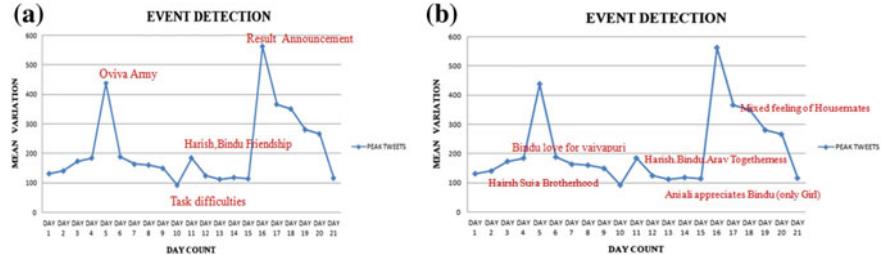


Fig. 3 Peak event detected in Day 16 by applying peak finding algorithm for the event ‘BiggBoss’

Table 3 Computation of bin-wise mean (M), mean deviation (MD) and difference (D) for tweets collected from Day 1 to Day 21

Day	M	MD	D	Day	M	MD	D	Day	M	MD	D
1	131	9		8	177	200	20	15	148	153	34
2	132	44	9	9	173	178	28	16	143	326	419
3	137	87	41	10	163	167	82	17	195	306	171
4	143	127	47	11	164	193	22	18	216	284	134
5	179	257	96	12	159	173	40	19	232	254	48
6	180	226	19	13	153	157	47	20	241	225	28
7	178	199	17	14	149	171	35	21	225	213	125

The content in bold indicates the peak event detected on Day 16 using the proposed approach

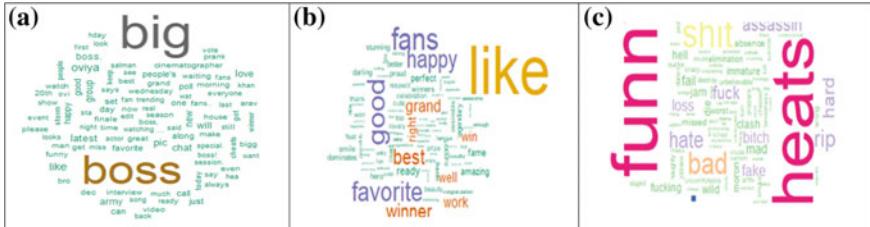


Fig. 4 Overall bag-of-words (a) generated from the preprocessed tweets for ‘BiggBoss.’ The positive (b) and negative (c) words that were used most frequently by the users

The n -gram approach is selected for better feature selection in text classification. Many research works proved that unigrams are better in feature selection, whereas the bigrams and trigrams are implemented to work on the contextual feature selection and statistical analysis. Table 4 explains the process of assigning scores for the text features based on the positive and negative word to arrive at a sentiment associated with a particular tweet.

The later part in hybrid sentiment analysis is applying Naïve Bayes classifier for achieving the truest sentiment prediction. From the 5290 tweets, n -gram classifier found 1817 tweets to be neutral. Nearly 34% of the total tweets were predicted to be

Table 4 Sample tweet represented in various n -grams namely unigram, bigram, and trigram with polarity score assigned for each term and the predicted sentiment of the tweet

N -grams	Tweets	Explanation	Sentiment
Uni-gram	I Don't Smoke In Bigg Boss	I / Don't / Smoke / In / Bigg / 0 -1 -1 0 0 Boss 0	Negative
Bi-gram	I Don't Smoke In Bigg Boss	I Don't - 0 Don't Smoke - +1 Smoke In - +1 In Bigg - 0 Bigg Boss - 0	Positive
Tri-gram	I Don't Smoke In Bigg Boss	I Don't Smoke - +1 In Bigg Boss - 0	Positive

Table 5 Classification result of sentiment analysis using Naïve Bayes classifier performed on the neutral tweets predicted by n -gram classifier

Prediction	Tweets			Sentiment
	Negative	Positive	Neutral	
Negative	166	70	0	
Neutral	0	0	27	
Positive	121	1433	0	

neutral for which there are possibilities to have either a positive or negative sentiment. Therefore, the Naïve Bayes classifier is applied to 1817 tweets, and the results are set forth in Table 5. The accuracy achieved after Naïve Bayes classifier was 89% which is appreciably high when compared to the unigram classifier.

5 Conclusion

Twitter is the highly preferred social media platform for posting news, reviews, comments, and updates of any organization, political party, products, and many other personal and professional stuffs. Being such a critical source for information generator, it is necessary to identify the peak events that occur in various timelines and extract the sentiment of the event and the user which can be further considered for decision making. The peak finding algorithm proposed by the authors finds the events that occur in small and large timeframe depending on the tweet frequency and author engagement. This approach efficiently handles the limitations faced due to data sparsity and redundancy of re-tweets. The hybrid sentiment classification results revealed a better performance by aggregating the prediction results of n -gram with bag-of-words model and Naïve Bayes classifiers. The word vectors or features considered by the hybrid sentiment classification are ranked for assigning the topic of the peak event. Therefore, the sentiment of the peak event detected can also be predicted for future events and user behavior.

References

1. Becker, H., Iter, D., Naaman, M., Gravano, L.: Identifying content for planned events across social media sites. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM). ACM, New York, NY (2012)
2. Aiello, L.M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., Goker, A., Kompatsiaris, I., Jaimes, A.: Sensing trending topics in twitter. IEEE Trans. Multimedia (2013)
3. Sayyadi, H., Hurst, M., Maykov, A.: In: Adar, E., Hurst, M., Finin, T., Glance, N.S., Nicolov, N., Tseng, B.L. (eds.) ICWSM: event detection and tracking in social streams. The AAAI Press (2009)
4. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based landmark and event detection on tagged photo collections. IEEE Multimedia Mag. **18**, 5263 (2011)
5. Schifferes, S., Newman, N., Thurman, N., Corney, D., Goker, A.S., Martin, C.: Identifying and verifying news through social media: developing a user-centred tool for professional journalists. Digit. J. **2**(3), 406–418 (2014)
6. Diakopoulos, N., Naaman, M., Kivran-Swaine, F.: Diamonds in the rough: social media visual analytics for journalistic inquiry. In: VAST (2010)
7. Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., Chi, E.H.: Eddi: interactive topic-based browsing of social status streams. In: UIST'10, p. 5 (2010)
8. Lu, Y., Tsaparas, P., Ntoulas, A., Polanyi, L.: Exploiting social context for review quality prediction. In: Proceedings of the 19th International Conference on World Wide Web WWW'10 (2011)
9. Boyd, D., Golder, S., Lotan, G.: Tweet, tweet, retweet: conversational aspects of retweeting on twitter. In: HICSS. IEEE (2010)
10. Dörk, M., Gruen, D., Williamson, C., Carpendale, S.: A visual backchannel for large-scale events. IEEE Trans. Vis. Comput. Graph. **16**, 1129–1138 (2009)

Land Use Land Cover Change Detection Through GIS and Unsupervised Learning Technique



Govind Kulkarni, Aniket Muley, Nilesh Deshmukh and Parag Bhalchandra

Abstract The remote sensing technology provides the means of classification of land cover with diversity of additional endless environmental variables over large spatial and moderate temporal extents. To investigate the land cover classification, remote sensing is useful as it provides a synoptic view with high level of information. Natural phenomenon and human intervention are major causes in land cover change can be easily seen with the help of satellite. Using land use land cover analysis methods, urban planner and policy maker can easily identify the change happened in some specific time period. In this study, unsupervised K-means clustering algorithm has been applied to investigate the land use land cover change in the time span of 2011–2016.

Keywords Unsupervised classification · K-means clustering · LULC · Rate of change

1 Introduction

Nowadays, the raising demands of fundamental human requirements and well-being, the land use land cover (LULC) becomes fundamental constituent for management

G. Kulkarni (✉) · N. Deshmukh · P. Bhalchandra

School of Computational Sciences, Swami Ramanand Teerth Marathwada University,

Nanded 431606, Maharashtra, India

e-mail: govindcoolkarni@gmail.com

N. Deshmukh

e-mail: nileshkd@yahoo.com

P. Bhalchandra

e-mail: srtmun.parag@gmail.com

A. Muley

School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University,

Nanded 431606, Maharashtra, India

e-mail: aniket.muley@gmail.com

of natural resources. Several reasons are responsible for these land cover changes which may make an effect on natural environment. Several studies have reported that the urban area must meet the fundamental requirement of growing economics. This can make an impact on environmental resources [1]. Urban monitoring is a vast challenging process and often performed as per the local government requirement to examine the urban trajectories, land use features. Several studies have reported that the urban area must meet the fundamental requirement of growing economics [1, 2, 5, 6, 11]. Expansion of urban area makes an impact on environmental resources. The use of LULC analysis method with their possibilities is important for lands selection, planning, and various schemes implementation to congregate the human requirements and welfare demands [13]. Nowadays, LULC is becoming fundamental constituent in current strategies for environmental monitoring change and natural resources management. The use of satellite observation is useful in managing natural resources with rapid and unrecorded environmental changes.

The study of LULC must be in continuous which permits long-term monitoring of earth's land observation with spatial details. This study allows studying geographic area with different time series. This involves the comparison of land used with different time of satellite images. The use of satellite observation is useful in managing natural resources with rapid and unrecorded environmental changes. Due to unsuccessful land uses planning in urban environment physical changes are happened with poor availability of spatial information. Further this affects global environment with green space minimization. The concept of LULC can be a game maker for policy developer. One can able to understand and can map the LULC change of a region [3]. In urban environment, knowledge about land resources with spatial distribution is essential which characterized by the well-organized plan, development activities done by people to produce the land cover type with change and maintenance. The change in earth surface is obtained by the livelihood activities which make unheard-of changes in ecosystem and environmental processes. The changed global scenario needs to be studied and analyzed which will provide the essential input to make decision of ecological and environmental planning management for future. While understanding the nature of the land, it is essential to differentiate the land in various classes. The land cover represents the physical condition of surface, forest, grasslands, etc. Further, we can also define land cover as the biophysical coverage of land like natural areas, forests, buildings, lakes, and roads. Cover defines the resources/features present on the surface of the earth. Whereas the land use defined as socioeconomic use of land for multipurpose environment which involved the human activities such as industrial zones, residential zones, etc. Other features such as natural, semi-natural, managed, or man-made are also involved in land cover type [8]. Information about LULC is one of the leading eras which provide the fundamental information about estimation rate, deforestation, habitat fragmentation, urbanization, wetland and soil deprivation and many other landscape-level phenomena. In general, human modification on earth surface refers to be LULC change. For the sustainable management of earth surface land use land cover studies are important to identify the dynamic changes during some specific time. Basically the LULC studies are driven by phenomena and anthropogenic activity which is sustainable development debate

in global scenario. Land prediction activities can be easily performed by land use land cover [1].

Reis [8] applied supervised classification technique on remotely sensed Landsat image of year 1976 and 2000, respectively, with GIS on Rize, northeast Turkey area. Six bands of satellite image were used with maximum likelihood algorithm for ground truth the change detection operation was performed on pixel for land cover changes. The zGIS functions were used for the analysis of topographic structure. Varadharajan et al. [11] used ASTER Satellite imagery with open source remote sensing software for the analysis of LULC of Coimbatore North Taluk. Their study was carried in between 1988 and 2011. ASTER Satellite imagery with 15 m spatial resolution was used. The prediction of land use changes is done with the help of GIS and image processing software. Suribabu [10] identified the urban growth of Tiruchirappalli city, Tamil Nadu with the help of Indian remote sensing satellite of 1989, 1992, 1995, 1998, 2001, 2004, 2007, and 2010 years. Supervised classification method was applied for classification of urban LULC. For verification and rectification of images CARTOSTAT-1 imagery of year 2009 with 2.5 m resolution was used. A minimum distance supervised classification algorithm was applied on training area with spatial pixel correction. Mirkatouli [7] used post-classification method to identify the LULC changes during the 1991 and 2013 years. TM and ETM images were used for the change detection analysis purpose. Hassan [2] investigated the exploration of urban spatial signature which explains the spatiotemporal LULC patterns in the five cities of Bangladesh. Landsat MSS with 60 m resolution of year 1973, TM 30 m (1989), ETM 30 m (2000), and OLI-TIRS 30 m (2014) satellite image of January and March of the respective years were used. Maximum likelihood classification (MLC) algorithm was used to categorize the land cover classes. Song and Deng [9] examine the change in ecosystem services through LULC change detection analysis of year 1988, 2000, and 2008, respectively. Value transfer method was used with consequent ESV. An electrical indicator was designed to assess the response in ESV change relative to LUCCs. Malmir et al. [5] investigated the change occurred in LULC of Ahwaz County in southwestern Iran during period of 26 years, from 1987 to 2013. Both unsupervised and supervised algorithms, i.e., ISODATA and Maximum Likelihood were applied for the classification. The overall accuracy was calculated by Kappa coefficient and the overall accuracy methods and finally, the LULC changes were identified using the cross tabulation method. Masta and Muringaniza [6] aim to find out the land use and the changes occurred in current Shurugwi district by applying the GIS and remote sensing techniques. The change detection analysis was based on three different years, i.e., 1991, 2000, and 2009. Classification method and image processing methods were applied. Image analysis operation was performed and change detection technique was used along with Google earth image to identify the change. In order to maintain the natural vegetation cover in urban region, it is important to examine whether the vegetation cover is reduced in the region or not? If yes how much amount of vegetation cover loss has been done during the specific time? At what rate the vegetation cover changed in specific time period? Focusing on the above scenario, we investigated the significance of change in LULC analysis during the time span of 2011–2016.

2 Material and Methodology

2.1 Study Area

Nanded city is major part of Marathwada region with multifunctional environment located on the bank of Godavari River. Nanded is divided into two parts, i.e., north and south Nanded. Nanded-Waghala Municipal Corporation is situated at latitude of $19^{\circ}15'N$ and longitude of $77^{\circ}30'E$ which is district headquarter. The current investigation is founded on the north Nanded which having six different regions as Anand Nagar, Taroda Budruk, Taroda Khurd, Shivaji Nagar, Vazirabad, and Itwara [4].

2.2 Remote Sensing Image

The purpose behind to use remote sensing data is to analyze and interpret data to verify the information extracted from remote sensing data. The referenced data is obtained from Google earth, United States Geological Survey (USGS), and Municipal Corporation of Nanded. Google earth image was used for rectifying the geographical boundary of the study area. The United States Geological Survey (USGS) provided the Landsat 5 and Landsat 8 images. Before investigating LULC operations, some basic operations like image preprocessing, subsetting of study area images, geometric correction, image classification, and image post classification were performed. One of the major objectives of this study is to identify the expansion of urban area growth with vegetation index values. The projection of these images consists of Universal Transverse Mercator (UTM) Zone 43 N, World Geodetic System (WGS) 84. For false color composition, this study used band combination of 4 3 2 (NIR, Red, Green) in Landsat 5 and Landsat 8 we used band combination of 5 4 3(Near-infrared, Red, Green). The study region data is extracted from Landsat imaginary and layer stacking of bands operation was performed and interested AOI was clipped from whole image. To investigate the land use land cover analysis, we used three land cover classes as other class, vegetation class, and urban area class. The re-codification of class may consist of several criteria such as group of intervals, geographical boundaries of study region. With the help of re-codification procedure, it is easy to simplify the raster dataset by making groups of similar cell or by identifying the cell values to particular class and removing it from the analysis. The reclassification process required prior knowledge about the study area. Unsupervised classification is a well-known classification technique in remote sensing. Unsupervised classification is commonly known as a clustering technique and is an efficacious process of remote sensing image partition data in multispectral feature space for extracting land cover information. The algorithm uses a group of pixels which having similar spectral features into unique cluster with some statistical defined criteria. In this classification technique, we do not need to specify the information about image feature. In this large number of unknown dataset is examined rather comparing the points to be classified with

training data in supervised learning. This technique required minimal amount of initial input from user. This process performs the natural grouping operation on image by selecting group of pixel to form spectral classes. The unsupervised classification technique groups the pixels into clusters. Once we have selected the clustering algorithm, we are able to identify the group of class that we want to generate. The class generation depends upon our interest [12]. After the cluster creation, we can manually identify each cluster and can merge the number of clusters as per our interest. In this investigation, we applied k-means clustering method as our study area is having small region. The generalized K-means clustering algorithm involves the following steps:

K-means clustering Algorithm

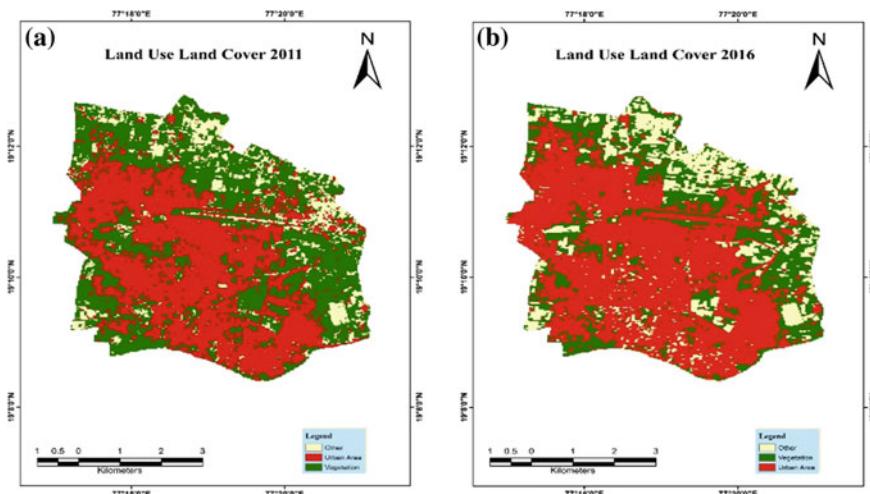
- S1: Randomly opening core cluster.
- S2: Compute the Euclidean distance of the feature vector from the cluster for every other cluster.
- S3: Assign the pixel to that cluster with minimum distance from its feature vector.
- S4: Compute the pixel mean of the feature vectors belongs to the cluster and renew the cluster.
- S5: If the change in the updated two consecutive clusters is less than a specified value, then stop the process. Otherwise go to S2.

3 Result and Discussion

Land use/land cover analysis of study region is assessed with the help of unsupervised classification for 2011 and 2016. The overall investigation area consists of 41,126,400 m² (41.1264 km²) land region. The investigation reveals that in 2011 the area occupied by urban area class was 18,448,200 m² (18.4482 km²), i.e., 44.85%. Area followed by vegetation class in 2011 was 18,333,900 m² (18.3339 km²), i.e., 44.57%. In 2011, the overall area occupied by other class was measured with 4,344,300 m² (4.3443 km²), i.e., 10.56% of land. During this investigation, it is also observed that there exists healthy balance in urban area and vegetation class. Most of the vegetation cover has been identified in east, northeast, northwest, and southwest region. The urban area represents the traditional part of the city with high density which is used for residential, commercial, public, and semi-public as the flow of urban region is toward the southeast to northwest region. The land cover class other was identified in northeast and north region significantly. Table 1 represents the area statistics of different land cover classes of year 2011. Figure 1a represents the land use land cover map of year 2011 with respective given classes.

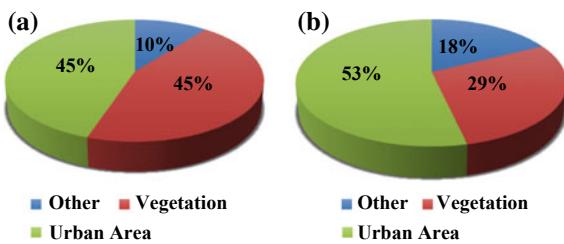
Table 1 Area Statistics

Sr. No.	2011			2016	
	Class	m ²	Percentage	m ²	Percentage
1	Other	4,344,300	10.56329	7,233,300	17.58797269
2	Vegetation	18,333,900	44.57939	11,872,800	28.87904762
3	Urban area	18,448,200	44.85732	22,020,300	53.54297969
	Total	41,126,400	100	41,126,400	100

**Fig. 1** **a** Land use land cover 2011. **b** Land use land cover 2016

For land use/land cover analysis of 2016, we observed that area occupied by urban area class was 22,020,300 m² (22.02 km²), i.e., 53.54% of area is covered by urban class. 28.86% of area is covered by vegetation class with 11,872,800 m² (11.87 km²) of land which is highly reduced in the study region. In the corresponding year, the overall area occupied by other class was measured with 7,233,300 m² (7.23 km²), i.e. 17.58% of land. As compared with the year 2011, significant growth has been identified in urban area and other land cover class. The uncontrolled development activities in the region were identified east region, west region and northwest region respectively which increased the urban area and mainly responsible for the loss of natural vegetation. Positive change has been identified in other class which expanded in the north and northeast region in large amount. A negative growth has been identified in the vegetation class with positive growth in urban area and other land cover class. The increasing population and development activities reduced the vegetation cover and vegetation pattern during the five years of time span. The unhealthy balance has been identified in natural vegetation and urban area. The great loss of vegetation cover was identified in the central region as development activities are increased. Table 1 reveals the area statistics of year 2011 and 2016 with respective land cover

Fig. 2 **a** Pie chart of year 2011 LULC. **b** Pie chart of year 2016 LULC



classes. Figure 1a–b represents land covered by these classes in respective year. The land occupied by vegetation class in the year 2011 is mostly converted in other class.

Further, the rate of change of these land cover classes is investigated which reveals the amount of change has been occurred in the respective classes during the time period of five years. The pie chart (See Fig. 2a–b) represents the overall percentage of respective land use land cover class of both the years 2011 and 2016. It is observed that, in 2011 the 45% of land was covered by urban area with 45% of vegetation class. 10% of land was covered by other class in year 2011. Additionally, in 2016 positive change has been identified in the urban area and other class as they occupied 53 and 18% of land in study region, respectively. Significant negative change has been identified in vegetation class as vegetation class was measured 29% of overall land. The overall scenario of the land use land cover analysis was represented in the form of digital maps and graphs (Figs. 1 and 2).

The rate of change reveals the change occurred in the respective class during the given time period and is investigated by following mathematical equation as:

$$\text{Rate of change} = \left\{ \frac{(Ae - Ab)}{Ab} \right\} * \left(\frac{1}{\Delta t} \right) * 100 \quad (1)$$

where Ae is the end year, i.e., (2016); Ab represents the base year (2011); and Δt stands for the time span (2011–2016), i.e., 5 years.

The obtained result reveals that, the significant positive change is measured in other class with 13.3% of change. In urban area, 3.87% of positive change is identified during the 5 years time period. The loss in vegetation class land is observed to be 7.04%. Highest rate of change is observed in the other class as most of the vegetation area is converted into other class. Large amount of agricultural land is converted in other class from north region and northeast of the study area. The remotely sensed data should be derived at local, regional, and global scales which explore the information briefly. In order to make any decisions, the remotely sensed information must have to be accurate. The obtained information is useful in scientific investigations and policy decisions. While deriving the information, errors always occur. The information derived through remotely sensed data is subjected to accuracy assessment (Table 2). To perform the accuracy assessment, random samples were collected from defined land classes. This study consists of three land classes; initially, 68 random points were selected for the year 2011 and 2016, respectively. Out of that 54 points were

Table 2 Accuracy assessment report

	Year 2011			Year 2016		
	Other	Vegetation	Urban area	Other	Vegetation	Urban area
Total Reference	1	35	32	14	50	4
Total Classified	1	31	36	6	58	4
Correct Number	1	26	27	5	48	3
Producer's accuracy (%)	100.00	74.29	84.38	35.71	96.00	75.00
User's accuracy (%)	100.00	83.87	75.00	83.33	82.76	75.00

correctly classified for the year 2011; and 56 points were correctly classified for year 2016, respectively. Further, accuracy was calculated with overall accuracy and kappa statistics for the completion of this process. The overall accuracy is based on the correctly classified point and total number of random points. For the year 2011, the overall classification accuracy was measured as 79.41% with kappa statistics of 0.6010. Moreover, for year 2016 the overall classification accuracy was measured as 82.35% with 0.4975 of kappa statistics.

4 Conclusion

In this study, Landsat 5 and Landsat 8 remote sensing images were used for years 2011 and 2016, respectively. Unsupervised K-means clustering algorithm was used for LULC operation to understand the geographical structure of study region. The result of this investigation reveals that significant change has been found in the study region during 2011–2016. It is observed that, the urban area growth 3.87% is due to the development activities. Further, result reveals that the land cover class, i.e. other has been increased with the 13% of growth. The land covered by vegetation class is decreased by 7.04% during the time span of 2011–2016. Also, uncontrolled development activities were identified in the east region, southwest, northwest region, and west region. Rapidly increasing development activities had reduced the physical environment which leads inappropriate land use changes. This study suggested that, in near future there is need to increase the natural vegetation cover in that area.

References

1. Ayele, Z.L.: Mapping Land Use/Land Cover Changes and its Effects on Urban Peri-Urban Agriculture in Debre Markos Town, Ethiopia, Doctoral Dissertation, Addis Ababa University (2017)

2. Hassan, M.M.: Monitoring land use/land cover change, urban growth dynamics and landscape pattern analysis in five fastest urbanized cities in Bangladesh. *Remote Sens. Appl. Soc. Environ.* **7**, 69–83 (2017)
3. Jing, X.: Modeling and Analyzing Land Use and Land Cover Change in Metropolitan Birmingham Area Using Landsat TM, OLI Sata. Doctoral dissertation, University of Alabama Libraries (2014)
4. Kulkarni, G.E., Muley, A.A., Deshmukh, N.K., Bhalchandra, P.U.: Multivariate statistical techniques for prediction of tree and shrub species plantation using soil parameters. *Model. Earth Syst. Environ.* **4**(1), 281–294 (2018)
5. Malmir, M., Zarkesh, M.M.K., Monavari, S.M., Jozi, S.A., Sharifi, E.: Urban development change detection based on multi-temporal satellite images as a fast tracking approach—a case study of Ahwaz County, Southwestern Iran. *Environ. Monit. Assess.* **187**(3), 108 (2015)
6. Matsa, M., Muringaniza, K.: An assessment of the land use and land cover changes in Shurugwi district Zimbabwe. *Ethiop. J. Environ. Stud. Manage.* **4**(1) (2011)
7. Mirkatouli, J., Hosseini, A., Neshat, A.: Analysis of land use and land cover spatial pattern based on Markov chains modelling. *City Territory Archit.* **2**(1), 4 (2015)
8. Reis, S.: Analyzing land use/land cover changes using remote sensing and GIS in Rize, North-East Turkey. *Sensors* **8**(10), 6188–6202 (2008)
9. Song, W., Deng, X.: Land-use/land-cover change and ecosystem service provision in China. *Sci. Total Environ.* **576**, 705–719 (2017)
10. Suribabu, C.R., Bhaskar, J., Neelakantan, T.R.: Land use/cover change detection of Tiruchirappalli City, India, using integrated remote sensing and GIS tools. *J. Indian Soc. Remote Sens.* **40**(4), 699–708 (2012)
11. Varadharajan, A., Kasinathap, P.: Assessment on Land Use Changes in Coimbatore North Taluk Using Image Processing and Geospatial Techniques (2012)
12. Venkateswaran, K., Kasthuri, N., Balakrishnan, K., Prakash, K.: Performance analysis of K-Means clustering for remotely sensed images. *Int. J. Comput. Appl.* **84**(12) (2013)
13. Wu, B., Yu, B., Yue, W., Shu, S., Tan, W., Hu, C., Liu, H.: A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data. *Remote Sens.* **5**(2), 584–611 (2013)

Innovating Digital Consumer Services in Wealth Management Ecosystem an Anthropomorphic Approach



Aditya Kulkarni and Swapnil Dambe

Abstract As consumer services become more and more dependent on technology, there is an increasing chance of missing out on the value benefits offered through traditional medium of human (in-person) connects. For a consultative consumer service, value outcomes (tangible as well as ones perceived by consumers) become key pillars upon which the business stands. Hence, while digitalizing a consultative consumer service, designing a business model that caters to human characteristics and values becomes a key factor. We applied Business Design thinking and Strategic Design and Management approach to transform a traditional Financial Advisory service to a digital one. Our basis was Service-tized product architecture with an interaction paradigm that imbues anthropomorphic characteristics. Anthropomorphism in software products has hitherto, primarily, been studied from a user interface and interaction perspective. However, we looked at human characteristics (physical, behavioral, and experiential) that have over the period become an essential part of the Wealth Management ecosystem. We show through our work how the derived anthropomorphic characteristics have created business value benefits to key stakeholders.

Keywords Anthropomorphism · Digital Advisory · Business Design · Strategic Design Management · Digital Consumer Service · Human Centered Experience · Wealth Management

A. Kulkarni (✉)

CGI Information Systems and Management Consultants Pvt. Ltd.,
Mumbai 400096, MH, India
e-mail: a.kulkarni@cgi.com; ads.sun.1987@gmail.com

S. Dambe

Vidyalankar Institute of Technology, Mumbai 400037, MH, India
e-mail: swapnil.dambe@gmail.com

1 Introduction

Any consumer service needs to reduce the various transaction costs that the consumer of the service would incur. Today's era is that of networked consumers and digital everything. Hence, with various means of connectivity available to digital service consumers, service expectations are high. This has affected almost every industry taken into consideration. One such industry affected is Wealth Management. Internet-based global connectivity has become a key discovery vehicle for one's wealth needs and a business model based on this that has reaped its benefits is Digital Advisory with its core engine driven by prediction algorithms.

Traditionally, Wealth Management has been a niche service with focus on consumers with large net investable assets. However, with advent of Digital Advisory, Economics of Wealth Management has changed. By increasing their target base from a few high-net-worth individual (HNWI) clients to mass affluent, the Wealth Management/ Financial Advisory firms can advise a larger group of consumers by expecting a lower average revenue per transaction and in return rely on associated economies of scale benefit.

Also, with the Internet, the primary transaction costs of search and retrieval of information for decision making have reduced to negligible for the service consumers. This has made information to be available at fingertips, and the expectation from financial advisors of the day is to do more of advisory than just selling a financial product.

Today, for successful digital transformation of a consultative service under Wealth Management, one needs to consider the consumer's heightened expectations from the service. These expectations have been developed through a trustworthy relationship developed by a human financial advisor, over the years, combined with the disruptive technological advances of the day.

1.1 A Service-tized Product Architecture

The digital product experience we strategized was primarily based on the premise of service-tized product architecture with the product staying true to the value propositions delivered by the traditional Financial Advisory service of in-person consulting, trust building through emotional connect, and long-term personal relationship building. We created a service-tized product architecture employing business design thinking approach.

Business design thinking helped to create a strong competitive advantage, business performance tangibility and was based on an approach of empathizing toward service consumers (Fig. 1).

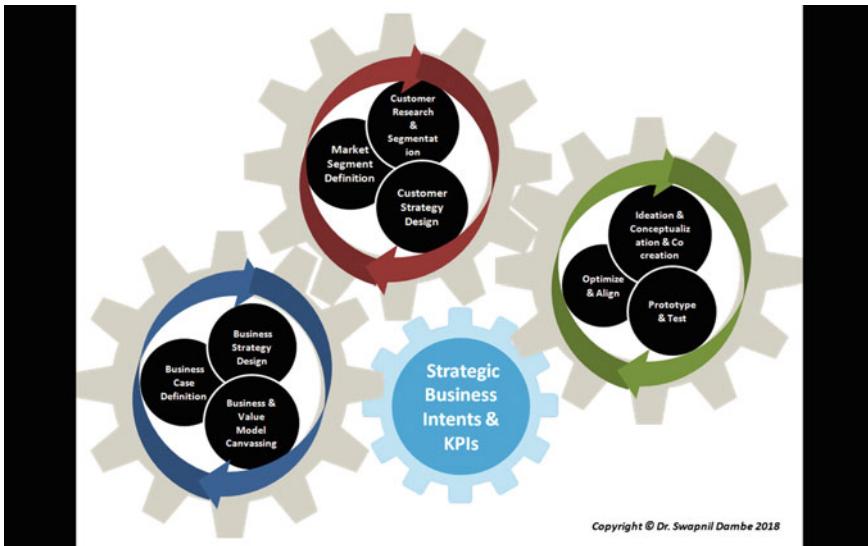


Fig. 1 Snapshot view of the business design thinking approach

1.2 *Employing Anthropomorphism by Studying ‘High-tech’ Versus ‘High-touch’ Patterns in Financial Advisory*

We introduced anthropomorphism as a concept to transform human characteristics into a digital solution so as to make the service consumers feel at ease while interacting with the Digital Advisory product. Emotional intelligence and human belief systems were studied to help the product to realize human emotions to some extent in certain real-life contexts/ situations and emulate a response based on it.

A Financial Advisory service is highly consultative in nature and is often times deeply rooted in a long-term relationship developed by financial advisor with client by understanding his/ her lifestyle preferences. As an anthropomorphic principle, we relate this to an understanding of human characteristics. Human characteristics play a fundamental role in designing one’s own lifestyle and paradigms. We have strategically conducted profitable target consumer research through segmentation and cultural understanding to meet with the segment’s needs to sustain and grow their wealth portfolios over a lifetime.

Conventional Digital Advisory product architecture leverages the technology benefits of digitalization and artificial intelligence, namely—virtualization, prediction algorithms, omni-channel digital connectivity that have a direct correlation with ‘high-tech’ patterns. These often fail when trying to emulate the patterns of in-person human advisory.

Table 1 Value benefits of ‘high-tech’ vs ‘high-touch’ patterns in Digital Advisory

‘High-tech’ Pattern Value Benefits	‘High-touch’ Pattern Value Benefits
Sophisticated risk-profiling questionnaires	Knowing client’s personal life, unique financial situations
Goal-based investment recommendations	Coaching the client’s to alleviate anxiety during unexpected market fluctuations
Auto-movement between different asset classes based on changing market conditions and individual investment needs	Addressing the last mile issue of convincing the client to make an investment

At the same time, decades of Wealth Management advisory experience has created styles and patterns adapted as best practices within the domain. These are primarily ‘high-touch’ in nature.

Successful digital transformation of Financial Advisory service would involve an amalgamation of ‘high-tech’ and ‘high-touch’ patterns. Key value benefits of some of these patterns are shown in Table 1.

2 Communication Gap in a Conventional Digital Service

Humans have a limited capacity to assimilate information, and if information is presented too rapidly and without adequate and timely response, this capacity exceeds and the two-way communication (between digital product and human) breaks down. This is what typically happens when a Digital Advisory service solely takes into consideration ‘high-tech’ patterns without considering the value benefits derived out of ‘high-touch’ patterns.

In such cases, prediction algorithms are able to generate portfolios and provide to the client, investment recommendations at quick successions; however, interaction between the digital product and human lacks much-needed impetus to build trust between the two to make a high-value judgment (like an investment decision). This is often mentioned as the ‘last mile issue of convincing the client.’

2.1 *Mapping Human Characteristics to Human Experience Satisfaction Levels*

Humans are characterized as self-centered, emotional, motivated, and intuitive by nature. We looked at key characteristic components of human beings and related it to the satisfaction-level expectation that is born out of it. This has helped us to understand the crux of an in-person advisor-client relationship, break it down into phases, and relate it to the various modules of the Digital Advisory product. Refer Table 2.

Table 2 Co-relation between human characteristic components and human experience levels of satisfaction

Human Characteristic Components	Human Experience Satisfaction Levels
Humans are intuitive—Most decisions are made based on mental shortcuts [1]	Mental satisfaction level is reached if humans are satisfied (satisfy + suffice) in a short time with the outcome
Humans are self-centered—View the world from their own perspective [1]	Rational satisfaction level is reached if the outcome of interaction matches with the perception developed by human from his/ her past experiences
Humans are emotional—Remember experiences based on how they are made to feel [1]	Emotional satisfaction level is reached through an emotional peak or closure experience
Humans are motivated—Driven by intrinsic motivation and are open to opportunities to satisfy it [1]	Motivational satisfaction level is reached if the interactive experience caters to an intrinsic need of humans

2.2 Understanding Client's (Human's) Social Engagement Style

Financial advisors have succeeded over the years by subconsciously (learned through work experience) catering to human characteristics. Financial advisors need to uncover communication styles suited to a specific client. Over a period, the advisor understands this and provides meaningful and personalized interactions.

Through domain best practices developed over the years, broadly three types of client engagement styles have been researched and identified [2, 3].

Open Engagement Style: Such an engagement style is interactive in nature with client displaying openness to get further insights and improvise on an existing idea.

Closed Engagement Style: Such an engagement style is more like a one-way conversation from the client-side intending for a quick closure.

Random Engagement Style: Such an engagement style might seem boundless and lacking a focus at first. Such clients cherish creative inputs and spontaneous expressions.

2.3 Understanding Human Decision-Making Patterns

Rational inattention deals with the effects of the cost of information acquisition on decision making [4, 5]. For example, when the information required for a decision is costly to acquire, decision makers may rationally take decisions based on incomplete information, rather than incurring the cost to get the complete information. This is the kind of decision making that most clients tend to rely upon in complex scenarios.

Thus, it becomes imperative to understand the complexity of decision making. Below is a classification of decision-making types. We have mapped these to experience satisfaction levels expected by clients. **Simple Decision making:** Applicable to problem solving that is Do-it-Yourself in nature. Such problems can ideally be solved through self-service solutions [6, 7]. Associated Experience Satisfaction benchmark—Mental Satisfaction level.

Complicated Decision making: Such decision making needs an external support or consultancy. Such could be in the form of tools to do the appropriate analysis and identifying the right solution [6, 7]. Associated Experience Satisfaction benchmark—Rational Satisfaction level.

Complex Decision making: A level of uncertainty is involved with such problems and its solution space. As there is no direct answer to such problems, tools that help in analysis are not enough. Advisor here plays the role of a counselor [6, 7]. Associated Experience Satisfaction benchmark—Rational and Mental Satisfaction levels.

Hyper-complex Decision making: As there is no single solution outcome known to such problems, there is a need to make use of simulation tools to help create prospective scenarios and their potential outcomes. Advisor here plays the role of a collaborator. [6, 7] Associated Experience Satisfaction benchmark - Rational, Emotional and Motivational Satisfaction levels.

3 Filling the Gap Using Anthropomorphic Principles

‘Anthropomorphism is imbuing the real or imagined behavior of nonhuman agents with human-like characteristics, motivations, intentions, or emotions’ [8].

In our case, we considered the intended consumer service, i.e., Digital Advisory as the nonhuman agent. As mentioned above, we looked at the core human characteristics, engagement styles developed through domain best practices and human decision-making patterns.

We began with creating a phase-wise breakdown of advisor’s in-person relationship with clients and mapped it with the generic phase that typically exists in a digital advisor (Table 3).

Table 3 Mapping Digital Advisory phase with break-down of advisor’s in-person relationship with clients

Breakdown of advisor’s in-person relationship with clients	Digital Advisory phase considered
Understand client’s personal life, unique financial situation	Onboarding
Probe client to find relevant information	Goal planning
Advise on taking right financial decision(s)	Customizing the Portfolio
Advisory at regular intervals	Investment dashboard, add new goals

With this, we then created a relational mapping of the digital advisor's considered phase with the proposed engagement style. Though we cannot cater to each individual client's engagement style, when clients interact with a digital advisor, a specific module within the product is expected to behave in one of these engagement styles for maintaining a human-like interaction.

For a particular digital advisory phase, the client must be engaged enough to make a valuable decision from his/ her end. Understanding various decision making patterns and creating a benchmark for a particular phase helps to associate it with suitable Key performance indicators (KPIs)/ success metrics.

4 Value Benefits of Anthropomorphized Digital Service

Engagement style and client decision-making pattern attained for a specific digital advisory phase was benchmarked at an expected client experience satisfaction level. In order for a Digital Advisory service to succeed it should attain the benchmark at the end of the specified phase. Success is measured in the form of business value attained during respective phases. Refer Tables 4 and 5 (Fig. 2).

Table 4 Deriving the key performance indicators (KPIs) for Digital Advisory product

Digital Advisory product phase	Advisor-client engagement style	Client decision-making pattern	Key Performance Indicator (KPI)
Onboarding	Closed—expect a cut to chase	Simple—need control and expect self-service	Being time bound and efficient
Goal planning	Open—need to vet ideas with other sources	Complicated to Complex—advisory-based solution	Solution needs to cover multiple possibilities and scenarios
Customizing Investment Portfolio	Closure-oriented—expect to reach their desired objectives	Complex—bounded flexibility in solution	Matching client's specific needs
Investment dashboard	Random—expect collaborative decision making	Hyper-complex—creativity in consulting	Value generated is appreciated over period of time

Table 5 Associated value benefits by attaining the expected client experience satisfaction-level benchmark KPIs

Digital Advisory product phase	Client experience satisfaction level	Key Performance Indicator (KPI)	Associated business value
Onboarding	Mental	Being time bound and efficient	Reduces early drop-offs and enables capturing valuable information for lead generation
Goal planning	Emotional	Solution needs to cover multiple possibilities and scenarios	Enough value received by client to remain engaged
Customizing investment portfolio	Mental, rational	Matching client's specific need	Service stickiness improves
Investment dashboard	Rational, emotional, motivational	Value generated is appreciated over period of time	Customer loyalty increases leading to brand advocacy

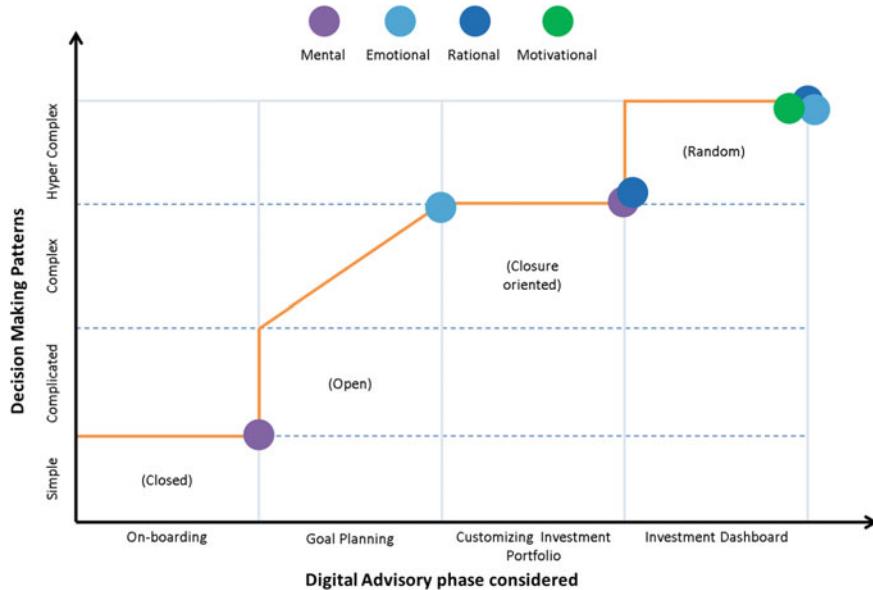


Fig. 2 Benchmarks to be attained for Digital Advisory phases

5 Conclusion

Conventional Digital Advisory as a business model made it viable for Wealth Management to expand its consumer base to the mass affluent. The key driver for this change has been technological advances that have given rise to ‘high-tech’ patterns.

However, for a conventional consultative service that is heavily dependent on patterns that are ‘high touch,’ we have identified human characteristics, client engagement styles and the decision-making patterns along with an expectation of a particular human experience satisfaction level to be attained. This plays a key role in engaging the service consumer over the long term. This, if not achieved, the product/consumer service is at risk of high consumer churn over a long-term period.

By achieving the experience satisfaction-level benchmarks, we are able to keep the clients engaged over a digital service channel similar to a long-term advisor-client relationship.

References

1. Temkin, B.: Humanize CX Putting People First, Temkin Group. <https://experiencematters.blog/humanize-cx-putting-people-first/>. Last accessed 2018/06/10
2. White, C., Richard, K.: Working with the Emotional Investor: Financial Psychology for Wealth Managers. Praeger, 8 Aug 2016
3. Think Advisor Article by Bob Clark, <https://www.thinkadvisor.com/2017/02/27/3-client-engagement-styles-advisors-need-to-unders/>. Last accessed 2018/06/10
4. Sallee, J.M.: Inattention, rational, efficiency, energy. *J. Law Econ.* **57**(3), 781–820 (2014)
5. Caplin, A., Dean, M.: Revealed preference, rational inattention and costly information acquisition. *Am. Econ. Rev.* **105**(7), 2183–2203 (2015)
6. Sullivan, P.J., Lazenby, D.L.: Scenario Selling: Technology and the Future of Professional Selling. Trafford Publishing, 15 July 2005
7. Kitces, M.: Kitces.com. <https://www.kitces.com/blog/real-time-collaborative-financial-planning-and-the-evolving-role-of-the-financial-advisor/>. Last accessed 2018/06/10
8. Epley, N., Waytz, A., Cacioppo, J.T.: On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* (US National Library of Medicine National Institutes of Health) **114**(4), 864–86 (2007)

Security Challenges and Solutions of IoT Ecosystem



Anshul Jain and Tanya Singh

Abstract The Internet of Things is the most innovative and focused technology being worked on these days. It is improving the standard of living of an individual and society. When a technology gets evolved, it brings along some shortcomings, which can always be targeted and exploited. The IoT challenges based on the basic security principles of confidentiality, integrity, and availability are addressed in this paper. An overview of the security constraints, requirements, mechanisms, and measures taken for the problems caused in secured communication of IoT ecosystem has also been talked about. The vulnerabilities of underlying IoT network are highlighted and various security challenges on different layers of IoT ecosystem have been discussed in this work. Different solutions are suggested for addressing the current challenges faced in IoT ecosystem, based on our study of existing vulnerabilities. Further added, it also briefs about different protocols for security in IoT.

Keywords Internet of things · Security · Privacy · Confidentiality · Integrity · Protocols

1 Introduction

IoT ecosystem is growing exponentially and is significantly impacting several domains. IoT applications are found both in industry and in academia. Industrial domain includes transportation and logistics, health care, society, aerospace, and aviation. Under academia analysis and researches are being performed on IoT applications. The application of IoT sensors is happening in all these sectors which is considered as one of the largest growth areas and is impacting everyone's life. A

A. Jain (✉)
AIIT, Amity University, Noida, UP, India
e-mail: anshuljain13@gmail.com

T. Singh
ASET, Amity University, Noida, UP, India
e-mail: tsingh2@amity.edu

huge number of sensors are used for everything from public health to safety to system monitoring or emission control [1]. The use of IoT has become an integral part of emerging medical and healthcare technologies, public administration and security. Smart technology is also being developed in almost all sectors around us. The information and data collected from several researches on the field are distributed and experiences are being shared through traditional as well as modern mechanisms. IoT sensors and technologies are being used and are contributing to the scientific cultivation with the increased quality. All these benefits bring some challenges along with them, for which solutions are required, which have been discussed in this paper.

The organization of this paper is as follows. Section 2 describes various security challenges faced by IoT ecosystem on different prospects. Section 3 explains about IoT security issues which includes confidentiality and access control. In Sect. 4, different complications of IoT layers are discussed. Section 5 suggests the solutions of the problems faced on different IoT layers. Section 6 explains security mechanisms in IoT system and points different IoT protocols. Finally, conclusion is presented in the last Sect. 7.

2 Challenges of Security

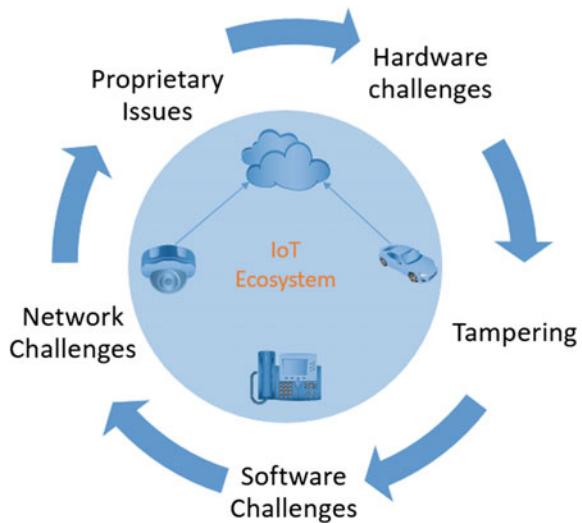
The three core issues with the IoT are confidentiality, integrity, and availability which impact both the human and business process. There are many interconnected and interacting components that communicate over public untrusted networks. These are combined and connected with the security, privacy, and open trust problems [2]. Security can be defined as the organized context consisting of principles, policies, procedures, techniques, and measures which are required to protect individual assets as well as the system. All the interactions must be secured by way of ensuring data and service provisioning of all the significant parties and restrict the amount of data, which helps in influencing the entire IoT. Figure 1 shows how different security challenges always revolve around our IoT ecosystem.

Following are the security challenges faced by IoT devices:

2.1 Proprietary Issues

- *Proprietary Applications*: Most of the IoT devices available in market are using proprietary applications which are not compatible with standard security solutions available in market.
- *Proprietary protocols*: Proprietary IoT devices also use proprietary protocols which are incompatible with existing network and security devices.
- *Interoperability*: Due to the devices using proprietary applications and protocols, their interoperability is almost impossible with other IoT devices, which causes scalability issues.

Fig. 1 Security issues around ecosystem of IoT



2.2 Hardware Challenges

- *Low Processing power*: IoT is deployed with very limited processing power due to multiple factors like cost. Due to which it is not able to process expensive algorithms.
- *Battery life*: Major chunk of IoT devices deployed in IoT network are battery guided, which is designed to run for a decade i.e. 10 years. To avoid battery drainage, IoT devices need to restrict execution of high energy consuming programs.
- *Limited Memory*: As compared to our latest digital devices, IoT devices are designed to run with a very low memory (RAM and Flash) which restricts them to directly run the latest security applications or programs on our IoT devices.

2.3 Tampering

IoT devices are deployed in remote areas; therefore, it is not possible to always monitor them. IoT devices should have tamper-resistant packing to keep themselves safe from malicious attackers who can steal them and modify or read their code. This can help them in doing reverse engineering on the device. Therefore, IoT devices should have tamper proof packing [2].

2.4 Software Challenges

- *Embedded software*: Software or operating systems which are embedded in IoT devices are very thin and lack security capability, robustness. Such devices are not designed to be fault tolerant.
- *Dynamic security update*: Security risks and challenges are increasing every day; therefore, IoT devices should be designed in such a way that they can get dynamic and real-time security updates, to save themselves from emerging security vulnerability.

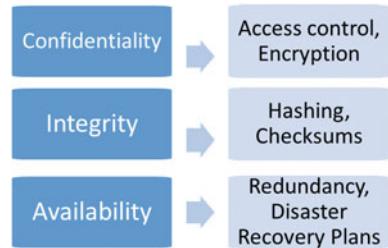
2.5 Network Challenges [3]

- *Mobile and Scalable*: IoT network is increasing exponentially, whereas traditional security solutions are not designed to handle this rate of growth and scalability, which makes these solutions incompatible for IoT network.
- *Bandwidth*: Exponential increase in network devices requires exponential bandwidth for communication and processing, currently which marks a limitation.
- *Diversity of IoT devices*: Today, there is a huge variety of IoT devices available which differs tremendously in their capability and processing power. Therefore, there are no security solutions available which can cater to such a wide gap.
- *Diversity in communication medium*: IoT devices can be connected using wide range of wireless and wired LAN and WAN. Therefore, a gateway is always required which makes their communication compatible with application.
- *Support for Multi-Protocol*: Due to its proprietary nature, IoT devices use a wide variety of network protocols for LAN communication. This makes this network limited with available security solutions.

3 IoT Security System

The key requirements of security system are Confidentiality, Integrity and Availability, it further manages access control and authentication, which creates concern in deployed IoT systems. To ensure secured communication in a data sharing environment, these requirements are necessary.

The block diagram of the IoT security system is presented in Fig. 2.

Fig. 2 IoT security system

3.1 Confidentiality and Integrity

As the approach presented in [4] which uses encapsulation mechanism such as smart business security, IoT application protocol also combines the communication with encryption, signature, and authentication to improve and establish a secured communication capability for IoT application development. The first fully implemented two-way authentication security scheme for the IoT based on the Internet standards is explained in [5]. The confidentiality and integrity have been analyzed in terms of key management systems applied on IoT. It is observed that key management systems can be classified into four major categories such as key pool framework, mathematical framework, public key framework, and negotiation framework. The KMS protocols are not suitable for IoT [6, 7].

There is no single IoT solution available which can guarantee a solution to basic security problems [8]. User authentication method as used in wireless network is suggested in [9], which can be used a context in IoT. This method uses key agreement scheme which enables a user to negotiate a session key between wireless network and an IoT device. This helps secure authentication within IoT ecosystem. This method aims at establishing authentication and access control using a session key based on best cryptography methods like AES. This method defines access control policies which are attribute-based and helps in boosting authentication among different nodes in IoT network [10].

3.2 Availability and Access Control

Proper redundancy and disaster recovery planning are required to ensure availability this can also be ensured using traditional BCP and DRP methods, whereas access control can be achieved by imposing restriction on a wide variety of users to access IoT network resources. Different roles can be based on rules, roles, or sensitivity of data which they are trying to access, z are assigned through access control. Two subjects identified to access data in [11] are data collectors and data holders. Every data access request according to policy and roles defined in the system must be verified and authenticated by the system, this will ensure that data is accessed only by

Layers	Technologies	IoT Components	Security Issues	Security Solutions	Security Target
Application Layer	Smart [City, Hospital, Office Homes, Grids], Intelligent Transport		<ul style="list-style-type: none"> - Software Bugs - Access Control(AAA) - Data Security - Cloud Security - Web Service Security - Fast & Secure Data Mining - Accurate Data Analytics - Middleware Security - Traditional Application layer attacks[DoS, DDoS, eavesdropping, etc..] 	<ul style="list-style-type: none"> - End to end encryption - IDS/IPS - Trust Management - User Authentication - User Authorization - PKI - Certification 	
	AI, Storage, Data mining, Cloud Computing, IoT Application				
	Bridge, Middleware				
Network Layer	IPv6, VPN, NGN, WiMAX, 3G, LTE, Wi-Fi, Bluetooth		<ul style="list-style-type: none"> - Network/LAN Security - Routing Issues - Bandwidth - Internet Security - Compatibility - Data on travel security - N no of nodes 	<ul style="list-style-type: none"> - Robust routing protocols - Network Access Control - IDS/IPS - Firewall - VPN access - Hop data encryption 	
Perception Layer	IoT Devices, RFID, GPS, ZigBee, Sensors, Actuators		<ul style="list-style-type: none"> - Physical device theft - RFID Security - Terminal Security - Data Sniffing 	<ul style="list-style-type: none"> - Tamper Proof - Chip Security(TPM) - Data encryption - Device Physical design - Sensor data security - PKI 	

Fig. 3 IoT layers and security solutions

authorized users according to the defined roles and not by anyone else [12]. Solution to the authentication problem of data streams which is outsourced is defined in CADS (Continuous Authentication on Data Streams) [13, 14].

Recently, a solution has been proposed to deal with access control of the heterogeneous devices [15] which suggest a subscriber method and a group membership scheme. This scheme works by adopting common confidential key denoted as group key [16] multicast communications can be secured. A general UML conceptual model is defined in which relationship between the entities involved and their IoT Infrastructure has been explained [17].

4 IoT Layers and Its Complications

The design of IoT ecosystem is complex and difficult, so is the IoT security problems. Therefore, it becomes necessary to identify all kinds of security problems existing in different layers [3, 24] and the possible potential attacks. Figure 3 briefs an overview of IoT ecosystem, it shows different IoT layers, technologies used therein, components, existing security issues, and security solutions for each layer. These attacks can lead to data leakage which can cause privacy and security issues for critical and sensitive data.

The security problems in each layer are explained below:

4.1 Perception Layer

It is the lowest layer in the IoT ecosystem and is responsible for communication using wired or wireless technologies like sensors, RFID, co-axial cable, etc. Most of these installations are at public places and many at remote locations which are easily accessible to attackers without any controls or checks, which makes it vulnerable to attacks.

Few attacks on perception layer can be classified as:

- *Physical attack*: Nodes can be physically harmed or tampered to get access to the network.
- *Capturing Nodes*: Radio keys can be captured using eavesdropping and thereafter gaining access into the node.
- *Side channel attack*: Attackers can try to access data using differential power analysis, time computation, or electromagnetic waves [18].
- *Replay attack*: Fabricated packet can be sent to obtain trust of the device and thereafter taking entry into the device.
- *Access fabrication*: Attackers can make use of race condition to fabricate access control and get into the device or network.

4.2 Network Layer

Second and middle layer in the IoT ecosystem is assigned to collect data from perception layer and pass it to upper layer for analysis and user access. This layer receives data from various types of sensors through IP and non-IP network. Below mentioned are few of the security issues we face on this layer.

- *Data theft*: All the data collected from different IoT devices gets processed on this layer and passed ahead, which is prone to attacks.
- *Compatibility issues*: IoT data is collected from wide range and type of sensor devices which comes both from IP and non-IP network, this can cause compatibility issues because security architecture is not yet designed to handle such scenarios.
- *DoS and DDoS*: Denial or distributed denial of service using ICMP or any other familiar ways can be used to disrupt service.
- *Historically Prevailing network problems*: Already prevailing network problems like eavesdropping and sniffing are also a major challenge to tackle with.

4.3 Application Layer

Top layer of IoT ecosystem faces totally different security issues, these are because of existing proprietary standards which makes every other application or device

incompatible with other [18, 19]. Today we do not have any universal standard which can make communication at this layer seamless. The application layer security is more complex, and few security problems such as data access permission, data protection, and recovery dealing with the mass data make it more vulnerable. Few of the security issues at this layer can be.

- *Software bugs*: There is no single way available wherein we can identify and eliminate all the bugs in our application. Attacks can always exploit zero-day vulnerabilities found in the software deployed at application layer.
- *IoT Service providers*: Most of the IoT services are using closed source, which makes it incompatible with existing security solutions like HIDS.
- *Data analytic or Artificial Intelligence*: Current AI services available are not as catchy as the AI word itself is. Successfully processing huge amount of data using data analytics is still a dream come true.
- *Data storage and protection*: We are yet to do a final estimate of the amount of data that is going to be generated out of the billions of connected IoT devices; therefore, storing, retaining, and providing security to data are a big challenge.

5 Suggested Solutions in IoT Layer

Diversity in IoT devices and network makes the application of security solutions complicated for any IoT ecosystem. There is no single security solution which can be unanimously applied to this heterogeneous network, therefore we need to apply different solutions at all layers. Below are the suggested solutions on respective layers [1].

5.1 Perception Layer

Sensors and radio frequency are two components that are deployed on this layer; therefore, main solutions are required on these two domains. Few of the suggested solutions are.

- *Chip Security*: Hardware security can be ensured using TPM (Trusted Platform Module), which will ensure secured communication.
- *Physical Security*: IoT device should have full metal shield to avoid tampering of devices deployed in public and remote areas.
- *Secure boot loader*: It should get deployed in OS chip; this will ensure that only secured software gets loaded after reboot.
- *Certification*: IoT hardware should be certified on defined standards by international organizations; this will ensure that tested hardware is deployed for critical operations.

- *Access control and Authentication:* Only authorized persons should be allowed to access IoT devices; this will ensure security.
- *Data encryption:* All data transmitted from device should be encrypted and key for decryption should be only with application server.
- *Sensor data Integrity:* Data generated by various sensors should be safely transmitted and processed, to ensure data is modified.
- *Natural Environment-specific device:* Various terrains have different natural uniqueness, extreme cold to extreme hot, therefore devices should be designed accordingly.

5.2 Network Layer

Network security solutions available for deployment are based on the existing network architecture, few as mentioned below.

- *IPSec security channel:* IPsec is a networking protocol which provides authentication and encryption. This protocol makes our communication secure and helps in preventing eavesdropping and tampering of data.
- *Firewall:* Firewall is used to restrict traffic based on rules defined on it. Now there are fifth generation firewalls available which can intelligently handle traffic.
- *VPN:* Virtual private network ensures that the communication happening on public network is secure and safe.
- *Secure routing protocol:* Routing and routed protocols used should be selected in a way to ensure security in data transmitted from device.
- *IDS/IPS:* IDS helps detect unauthorized access to network and send an alarm, where IPS behaves intelligently and restrict attacks.
- *Hop by hop data encryption:* Data transmitted on network should be encrypted to avoid sniffing or eavesdropping.
- *Access control and authentication at network layer:* Proper authentication and authorization should be enabled on this layer also, this will ensure that any unauthorized person does not have access to our network.

5.3 Application Layer

- *End to End Encryption:* In this communication only end users can read the messages communicated between two parties, it helps to eradicate attacks such as eavesdropping and wiretapping. E2EE is designed to defeat any attempt by third party to tamper data or restrict any sniffing on network traffic. All the data transmitted is encrypted which can only be decrypted by users having valid key.
- *IDS/IPS:* IDS helps detect unauthorized access to network and send an alarm, where IPS behaves intelligently and restricts attacks.

- *Trust management:* It can be ensured using PKI which enables unknown devices on public network to trust each other.

6 Security Mechanisms in IoT System

Specialized services for collecting data from IoT devices, data processing, and connection with other services for decision-making about further activities like activating actuators, etc. [19–21] are represented by IoT core network.

It is important to highlight that data is secured from others, but it should also be kept in mind that IoT represents the network of devices (“the social network of devices”) that exchange data, so the data should be available but in a modulated manner. Development of pervasive computing as well as the development of context aware computing [22–24] is to be allowed by huge data.

With the emerging growth of the IoT, it was possible that different protocols are used for communications between machines (M2M) [25]. Few limitations of protocols are:

- Communication medium with restricted bandwidth.
- Hardware devices with limited resources.
- Transmitting data through disparate channel.
- Huge participation.

Below protocols are used for Machine to Machine and IoT network [26]

- Advanced Message Queuing Protocol (AMQP) [27].
- Message Queuing Telemetry Transport [28].
- Constrained Application Protocol (CoAP) [29, 30].

7 Conclusion

Massive deployment of IoT has enabled it to enter in most of our critical daily operations like homes, health care, transportation, grid; therefore, induction of security in ecosystem of IoT becomes very crucial. Various security concerns and challenges which roam around IoT system are presented in this paper. Different facets of these challenges like hardware, software, and network are explained categorically. The comprehensive overview presented in this paper has several issues open for discussion; it also explains security system of IoT which is a challenge in today’s information technology network. Different possible solutions have also been classified according to different layers of IoT ecosystem, which gives an enhanced understanding of how to handle such challenges. A brief description of some protocols helpful in smooth functioning of IoT network is also provided in the paper. The solutions suggested by us will help researchers address a few security challenges in IoT ecosystem,

but still, there is a huge gap to be filled. Many researches have already taken place in past and few are ongoing which are addressing the common security challenges but still road does not end nearby. A lot more research must be done but we hope this paper can help in streamlining few of the challenges and our suggestions will help in unified and secured deployment of IoT network across the globe.

References

1. Ashibani, Y., Mahmoud, Q.: Cyber physical systems security: analysis, challenges and solutions. *Comput. Secur.* **68**, 81–97 (2017). <https://doi.org/10.1016/j.cose.2017.04.005>
2. Deogirikar, J., Vidhate, A.: Security attacks in IoT: a survey. In: International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017)
3. Zhao, K., Ge, L.: A survey on the internet of things security. In: 2013 Ninth International Conference on Computational Intelligence and Security, Leshan, pp. 663–667 (2013)
4. Kothmayr, T., Schmitt, C., Hu, W., Brunig, M., Carle, G.: Dtls based security and two-way authentication for the internet of things. *Ad Hoc Netw.* **11**(8), 2710–2723 (2013)
5. Roman, R., Alcaraz, C., Lopez, J., Sklavos, N.: Key management systems for sensor networks in the context of the internet of things. *Comput. Electrical Eng.* **37**(2), 147–159 (2011)
6. Du, W., Deng, J., Han, Y., Varshney, P., Katz, J., Khalili, A.: A pairwise key predistribution scheme for wireless sensor networks. *ACM Trans. Inf. Syst. Secure (TISSEC)* **8**(2), 228–258 (2005)
7. Piro, G., Boggia, G., Grieco, L.A.: A standard compliant security framework for ieee 802.15.4 networks. In: Proceedings of IEEE World Forum on Internet of Things (WF-IoT), Seoul, South Korea, pp. 27–30 (2014)
8. Turkano, M., Brumen, B., Hlbl, M.: A novel user authentication and key agreement scheme for heterogeneous ad hoc wireless sensor networks, based on the internet of things notion. *Ad Hoc Netw.* **20**, 96–112 (2014)
9. Ye, N., Zhu, Y., Wang, R.C.B., Malekian, R., Lin, Q.-M.: An efficient authentication and access control scheme for perception layer of internet of things. *Appl. Math. Inf. Sci.* **8**(4), 1617–1624 (2014)
10. Alcaide, E., Palomar, J., Montero-Castillo, A.: Ribagorda, anonymous authentication for privacy-preserving iot target driven applications. *Comput. Secure.* **37**, 111–112 (2013)
11. Hu, C., Zhang, J., Wen, Q.: An identity-based personal location system with protected privacy in IoT. In: Proceedings—2011 4th IEEE International Conference on Broadband Network and Multimedia Technology, IC-BNMT 2011, Shenzhen, China, pp. 192–195 (2011)
12. Papadopoulos, S., Yang, Y., Papadias, D.: Cads: continuous authentication on data streams, In: Proceedings of the 33rd International Conference on Very Large Data Bases, VLDB '07, Vienna, Austria, pp. 135–146 (2007)
13. Papadopoulos, S., Yang, Y., Papadias, D.: Continuous authentication on relational data streams. *VLDB J.* **19**(1), 161–180 (2010)
14. Cherkaoui, A., Bossuet, L., Seitz, L., Selander, G., Borgaonkar, R.: New paradigms for access control in constrained environments. In: 2014 9th International Symposium on Reconfigurable and Communication-Centric Systems-on-Chip (ReCoSoC), Montpellier, pp. 1–4 (2014)
15. Veltri, L., Cirani, S., Busanelli, S., Ferrari, G.: A novel batch-based group key management protocol applied to the internet of things. *Ad Hoc Netw.* **11**(8), 2724–2737 (2013)
16. Sicari, S., Rizzardi, A., Cappiello, C., Coen-Porisini, A.: An NFP model for internet of things applications. In: Proceedings of IEEE WiMob, Larnaca, Cyprus, pp. 164–171 (2014)
17. Zhang, Y., Lee, W.: Intrusion detection in wireless ad-hoc networks. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. ACM, pp. 275–283 (2000)

18. Verma, H., Chahal, K.: A review on security problems and measures of Internet of Things. In: 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, pp. 71–76 (2017)
19. Ivan, T., Dušan, B., Komlen, L., Miodrag, B., Ivana, O.: Security Mechanisms in IoT (2017)
20. Sharma, V., Understanding Constrained Application Protocol. Exilant Technologies Pvt (2014)
21. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
22. Alaba, F.A., Othman, M., Hashem, I.A.T., Alotaibi, F.: Internet of Things security: a survey. J. Netw. Comput. Appl. **88**, 10–28 (2017). ISSN 1084-8045. <https://doi.org/10.1016/j.jnca.2017.04.002>
23. Smailagic, A., Kogan, D.: Location sensing and privacy in a context-aware computing environment. IEEE Wirel. Commun. **9**(5), 10–17 (2002)
24. Rahman, R.A., Shah, B.: Security analysis of IoT protocols: a focus in CoAP. In: 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC), IEEE, pp. 1–7 (2016)
25. Chen, Y., Kunz, T.: Performance evaluation of IoT protocols under a constrained wireless access network. In: 2016 International Conference on Selected Topics in Mobile & Wireless Networking (Mowed), IEEE, pp. 1–7 (2016)
26. Zamfir, S., Balan, T., Iliescu, I., Sandu, F.: A security analysis on standard IoT protocols. In: 2016 International Conference on Applied and Theoretical Electricity (ICATE), IEEE, pp. 1–6 (2016)
27. Aijaz, A., Aghvami, A.H.: Cognitive machine-to-machine communications for Internet-of-Things: a protocol stack perspective. IEEE Internet of Things J. **2**(2), 103–112 (2015)
28. Freier, A., Karlton, P., Kocher, P.: The Secure Sockets Layer (SSL) Protocol Version 3.0 (2011)
29. Karagiannis, V., Chatzimisios, P., Vazquez-Gallego, F., Alonso-Zarate, J.: A survey on application layer protocols for the internet of things. Trans. IoT Cloud Comput. **3**(1), 11–17 (2015)
30. Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M., Ayyash, M.: Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutorials **17**(4), 2347–2376 (2015)

A Wireless Sensor System to Monitor Banana Growth Based on the Temperature



B. Geethanjali and B. L. Muralidhara

Abstract Banana is a wonderful commercial fruit with several health edges. Banana is additionally helpful in many ways in our day to day life. Banana is cultured in all districts of the world. Bananas are developed in tropical condition. The expansion and development of the crop are greatly influenced by factors such as temperature, humidness and soil wetness. A wireless sensor observance system is developed to determine temperature, humidness and soil wetness. The collected data is stored in the SQLite database in Raspberry Pi along with the time stamp, and the data is displayed on the local web host server. Temperature data is compared with the optimum temperature values using fuzzy rules. Based on the fuzzy rules, an alert is remitted to the farmer through email. This helps the farmer to keep track of the growth of the fruit and reduce the loss that happens because of low or high temperatures.

Keywords Chilling injury · Sunburn · Fuzzy rules · Sensors · Temperature

1 Introduction

Banana is one of all the antiquated fruits cultured by human from the ancient period. References to banana have been utilised in religious rites, oftentimes occur in religious writing. The bananas belong to the genus *Musa* of the family *Musaceae* [1]. Banana is a torrid plant grown in many regions of the world. In India, banana is considered as an auspicious plant and is a part of various customs and rituals. The banana plant is useful in many ways. Almost all the parts of the banana plant are edible and are a rich source of potassium, magnesium, vitamins and fibre. They are furthermore used in the preparation of ayurvedic medicines, arts and crafts. The crop is thought to be perceptibly at risk of the warm and cold weather conditions. The nearly suitable climate for the crop growth is one with the warm damp weather for the whole of

B. Geethanjali (✉) · B. L. Muralidhara
Department of Computer Science and Applications,
Bangalore University, Bangalore, India
e-mail: geethanjalibkris@gmail.com

the year while not sturdy winds. Favourable factors deciding its distribution are rainfall in way over 100 mm per month and a temperature range of 13–38 °C with relative humidness of 75–85% [2]. Growth and blooming of bananas are impaired by temperatures outside this range. Temperature, humidness and moisture conditions are important for seed germination and productivity of banana plants. Humidness influences plant diseases and post-harvest storage of banana. We have emerged with a system that focuses on the temperature, humidness and soil moisture which reinforces the banana growth. The factors temperature, humidness and soil moisture that are essential for the crop growth are determined. Once the normal range of any factors exceeds, the farmer is instantly notified.

2 Effects of Temperature, Humidness and Moisture on Banana Plantation

Several growth and development processes like leaf production, leaf area, leaf unfurling, ripening of fruit, bunch weight and so on are contrived with the prevailing temperatures at any given time. Transportation and storage conjointly want essential temperature to be maintained. The regular growth of the banana begins at 18 °C, attains ideal at 27 °C, then decreases and draws up at 38 °C [2]. Temperature affects the position of the leaves. The leaves grow upright and horizontal in below cool conditions. Temperatures below 10 °C cause delay in inflorescence and deformity of bunches. The speed of emergence of new leaves and rate of fruit growth are noted to be largely dominated by temperature. Chilling injury results from exposure of fruit to terribly low temperature for perilous times. Chilling might occur within the field wherever air temperature drops below 13 °C for many hours. Young bunches might fall out, the leaves might slowly die and exposure of the bunch to the sun could cause sunburn once the fruit flip black [1]. Chilling within the field will be prevented to some extent by covering the fruit with paper baggage lined with perforated polythene. A single night below freeze temperatures quickly kills the banana plant and can ruin the whole plantation. Bananas cannot be grown in areas of frequent frosts.

High air temperatures and splendid daylight cause sunburn of uncovered natural fruit, eminently on the most elevated hands of the bunch [3]. Sunburn is caused by excessive exposure of the bunch to the direct sunlight takes place when the temperature rises above 42 °C. The fingers seem dry and switch yellow ahead of time. In India, sunburn had been known as the most significant reason for injury in the banana plantation. Covering the plant with dried banana leaves has been recommended to scale back the consequences of sunburn.

Very cool winds in winter or hot dry winds in summer are well-known to attract each fruit quality and yield of banana. Sensible soil drainage is thought to be an important condition that favours banana growth. Water logging decreases the bunch weight and different growth parameters. The banana plants require a ceaseless well-spring of a lot of soil dampness for ideal development [4]. The primary signs of

water stress may be a yellow forged or pale green colour of the foliage, amid long periods of the transient weakening of the leaves. Leaf damage, retardation of leaf emission, short-fingered fruit and softening of fruit before reaching harvest stage are results of the effects of drought.

The humidness of air in storage rooms directly affects the quality of the fruit. If humidness is too low, wilt or shrivelling is likely to occur. High humidness is helpful for wound healing and periderm formation. High rates of 85–100% humidness are suggested in storage rooms to retard softening and wilt.

3 System Overview

Figure 1 shows the block diagram of the banana plant monitoring system. Raspberry Pi is the core of the system. Using MQTT, it subscribes to ESP8266 and receives temperature, humidness and moisture data. The data received is stored in the SQLite database. Fuzzy rules are used to compare the temperature value stored in SQLite, and based on the rules, an appropriate alert is sent to the farmer's email ID along with the appropriate action to be taken. The stored data is displayed dynamically on the local web host. The Python script is used for programming.

3.1 Equipment Implementation

To decide the estimations of temperature, humidness and soil dampness of the banana plant, we have built up a remote sensor framework. The proposed system collects the values of three main factors that affect the growth of banana plant. They are temperature, humidness and soil moisture.

Raspberry Pi. Raspberry Pi 3 is a UNIX working framework, and Quad-64-bit ARM-based microcomputer [5]. It is a master card sized board, and it operates within the same method as a customary computer, requiring a keyboard for command entry, a display unit and a power supply. It uses SD nonvolatile storage card for the filing system. Web connectivity is via 100 Mbps Ethernet and 802.11n wireless LAN. There is a row of forty general purpose input/output (GPIO) pins on the sting of the Raspberry board.

DHT11. DHT11 is a temperature and humidness detecting component, and it incorporates a resistive-type stickiness measure section, a negative temperature constant (NTC) temperature measure part and a superior 8-bit microcontroller. It provides graduated digital signal output.

Octopus Soil Moisture Sensor. Octopus soil moisture sensing element is employed to browse the quantity of wetness present in the soil. This sensing element uses two probes. Through the probes, current is passed to the soil, and then, the sensing element reads the resistance offered by the soil to urge the moisture

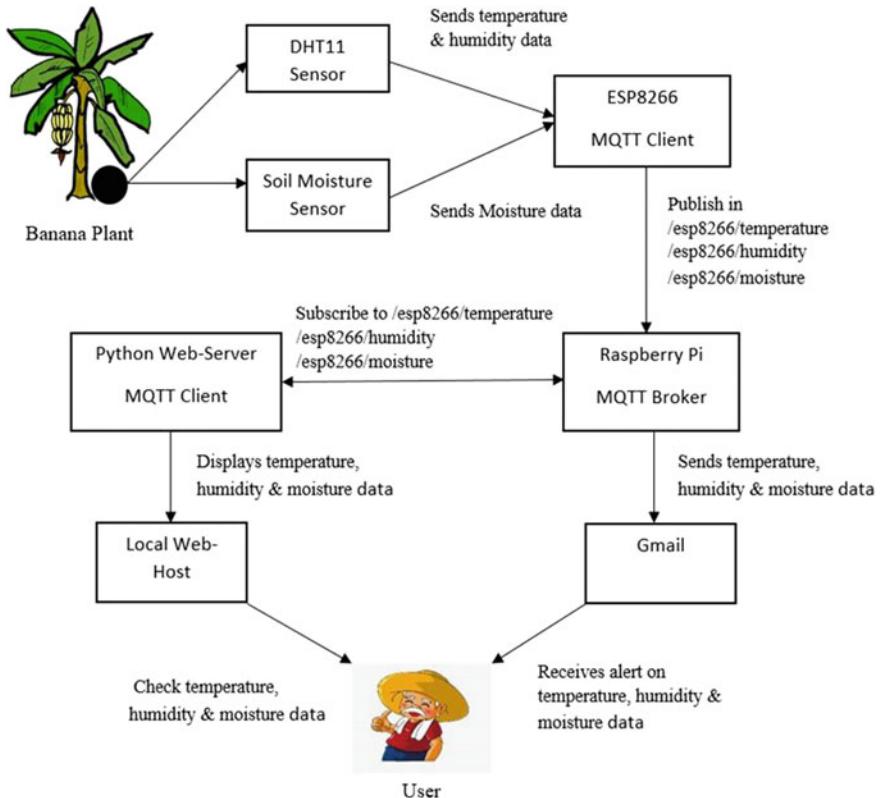


Fig. 1 Block diagram of banana plant monitoring system

level. Sodden soil conducts power effectively and offers less opposition, and dry soil conducts power ineffectively and offers a considerable measure of obstruction.

ESP8266. The ESP8266 wireless fidelity module is System on Chip (SoC) with integrated TCP/IP protocol stack which is able to give any microcontroller to access Wi-Fi network.

3.2 Working

DHT11 and soil moisture sensors are connected to ESP8266 and are programmed in a way to read the data from sensors for every 30 min. Rest of the time ESP8266 is set to sleep so that the sensors do not draw power from the ESP8266 module. This module is powered using a battery. The module is set up under the banana plant.

ESP8266 sends the information collected from the sensors to Raspberry Pi through Wi-Fi affiliation using message queuing telemetry transport (MQTT). It is a pub-

Table 1 Temperature effects on banana growth

Temperature (°C)	Effects	Control measures
<10	Chilling injury	Covering the fruit with paper bag lined with perforated polythene
18	Growth begins	Monitor irrigation
27	Optimum for development	Application of pesticides, digging and weeding
38	Growth declines	Monitor the fruit for harvest
>38	Sunburn/choking	Covering the plant with dried banana leaves, monitoring irrigation

lish/subscribe extremely basic and lightweight electronic informing convention, intended for obliged gadgets and low data transmission. Raspberry Pi subscribes to ESP8266 topics /esp8266/temperature, /esp8266/humidness and /esp8266/moisture using MQTT broker. ESP8266 publishes these topics to Raspberry Pi.

Temperature, humidity and soil moisture data are now stored in Raspberry Pi using the SQLite database. SQLite is chosen since the database needs no administration, and it works well in devices that need to operate without skilled human support. The tabulated data with the time stamp is displayed on the local web host using Python web server and is sent to farmer's email address using Secure Simple Mail Transfer Protocol (SSMTP). SSMTP program is used to forward email automatically from a system to an external email address, which is configured mail host (mail hub).

Fuzzy rules are created based on the data in Table 1 [6]. Based on these fuzzy rules, an email alert is remitted to the farmer's mail address.

3.3 Fuzzy Rules

For the temperature: the fuzzy partition is outlined by five fuzzy sets, $T = (C, G, O, D, S) = (\text{Chilling Injury}, \text{Growth}, \text{Optimum}, \text{Growth decline}, \text{Sunburn})$ is that the i th fuzzy subset, $i = \{1, 2, 3, 4, 5\}$. These subsets cover the fuzzy domain $x = [5^{\circ}\text{C}, 45^{\circ}\text{C}]$.

Fuzzy Rules developed are:

If temperature is <10, then alert chilling injury

If temperature is >10 and ≤ 18 , then growth has begun

If temperature is >27 and ≤ 30 , then alert optimum development stage

slno	date	time	temperature(c)	humidity(%)	soilmoisture
1	24-05-2018	11:45am	31	64	928
2	24-05-2018	12:15pm	31	64	928
3	24-05-2018	12:45pm	31	64	928
4	24-05-2018	1:15pm	31	64	928
5	24-05-2018	2:45pm	31	64	928
6	24-05-2018	3:15pm	31	64	928
7	24-05-2018	3:45pm	31	64	928
8	24-05-2018	4:15pm	30	64	928
9	24-05-2018	4:45pm	30	64	928
10	24-05-2018	5:15pm	28	64	928
11	25-05-2018	11:45am	29	62	948
12	25-05-2018	12:15pm	29	62	948
13	25-05-2018	12:45pm	29	62	948
14	25-05-2018	1:15pm	29	62	948
15	25-05-2018	2:45pm	29	62	948
16	25-05-2018	3:15pm	29	62	948
17	25-05-2018	3:45pm	29	62	948
18	25-05-2018	4:15pm	29	62	948
19	25-05-2018	4:45pm	28	62	948
20	25-05-2018	5:15pm	28	62	948
21	26-05-2018	11:45am	30	57	928
22	26-05-2018	12:15pm	30	57	928

Fig. 2 Sample temperature, humidness and soil moisture data tabulated in SQLite database

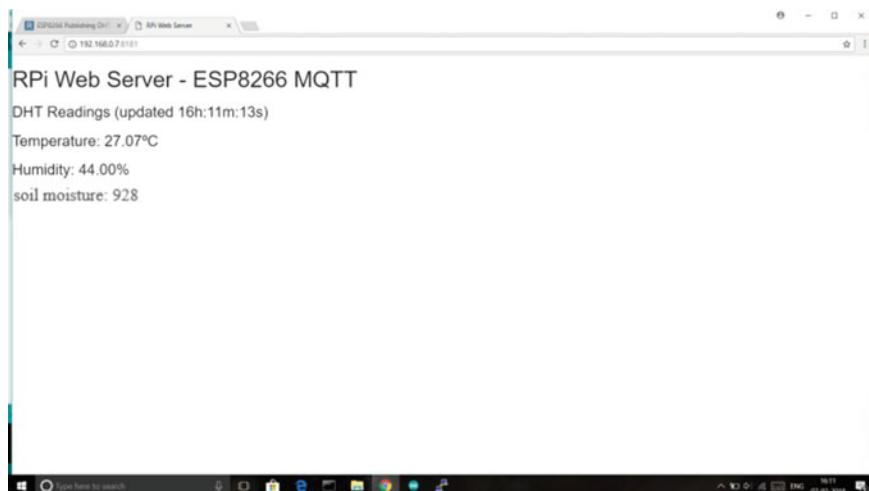
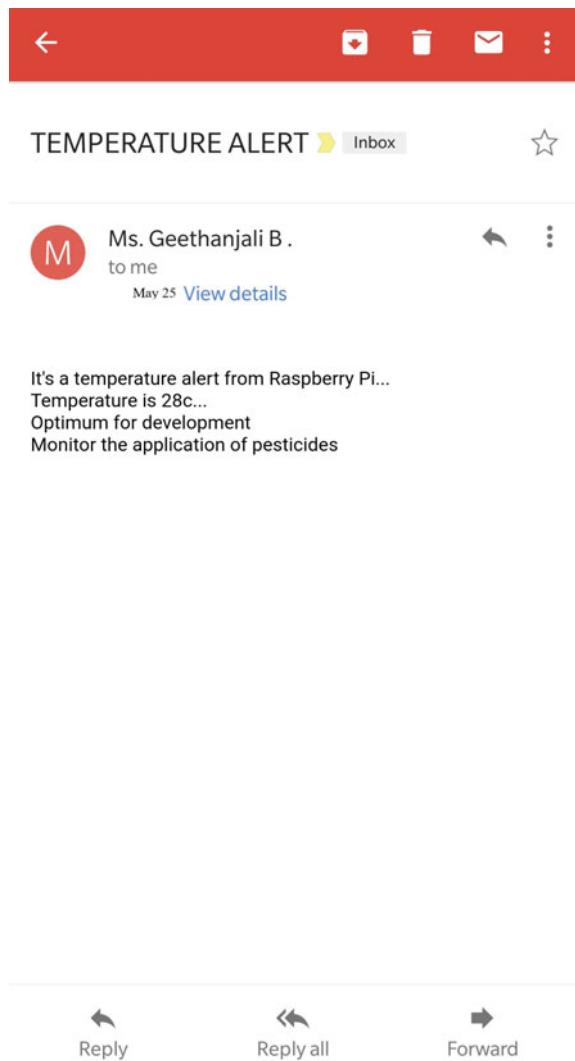


Fig. 3 Data displayed on local web host

Fig. 4 Sample email alert

If temperature is >35 , then alert growth declines
If temperature is >38 , then alert sunburn.

4 Conclusion

The banana plantation taking twelve to eighteen months' time from planting to reap comes beneath the influence of a spread of seasonal conditions at totally different

growth stages, resulting in reduced leaf production or scrawny growth in winter because of cool temperatures and premature ripening of fruit because of hot summer weather. The behaviour of banana plant growth with respect to climatic conditions makes it necessary to find factors such as temperature, humidity and soil moisture. These factors impact the advancement of the yield. The Wireless sensor system is used to collect temperature, humidity and soil moisture data. Figure 2 shows the Sample of collected data that is tabulated in SQLite database. Figure 3 shows the data displayed on local web host. The collected temperature data is compared with its optimum value, necessary for the growth of the plant. Using fuzzy rules on compared values of temperature data, a precise mail alert is sent to the farmers. Figure 4 shows a sample email alert. As a result farmers can take needful actions required for the growth of the plant and avoid loss. The collected data about these factors is also useful in further studies about the crop. The research exposed to view that very meagre quantitative information is out there on the banana climate relationship. Quantitative relations remain to be found out between weather factors and banana growth and development.

References

1. Sastry, P.S.N.: Agrometeorology of the Banana Crop. World Meteorological Organization, Geneva (1988)
2. Banana—Vikaspedia. <http://vikaspedia.in/agriculture/crop-production/package-of-practices/fruits-1/banana>
3. Deuter, P., White, N., Putland, D.: Critical Temperature Thresholds Case Study Banana. Agri-Science Queensland, DEEDI; Growcom, Queensland
4. Abd El-Kader, A.M., Saleh, M.M.S., Ali, M.A.: Effect of soil moisture levels and some antitranspirants on vegetative growth, leaf mineral content, yield and fruit quality of Williams Banana Plants. *J. Appl. Sci. Res.* **2**(12), 1248–1255 INSNet Publication (2006)
5. Deshmukh, A.D., Shinde, U.B.: A low cost environment monitoring system using raspberry Pi and arduino with Zigbee. In: International Conference on Inventive Computation Technologies (ICICT), Coimbatore, pp. 1–6 (2016)
6. Abu, M.A., Yacob, M.Y.: Development and simulation of an agriculture control system using fuzzy logic method and visual basic environment. In: 2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems (ROBIONETICS) Yogyakarta, Indonesia, November 25–27 (2013)

User Review Classification and Star Rating Prediction by Sentimental Analysis and Machine Learning Classifiers



Aagam Shah, Komal Kothari, Umang Thakkar and Satvik Khara

Abstract With the digital media explosion, in today's increasing Internet usage, the data generated is wide and varied. Huge volumes of new data are injected daily into the Web for various prospects. Procedures like text mining and analysis are required to make the best use of this potential. User review analysis benefits us with the exact understanding of the user's feedback toward the product. In this paper, we have proposed a unique approach by performing abstract-level sentimental analysis of user review by n-gram classification and POS tagging. This classification is then used as entropy for machine learning algorithm. This paper leverages upon the proposed methodology with promising outcomes and improved accuracy by evaluating the data with the help of two algorithms, MaxEnt model and Naïve Bayes classifier, after analyzing few algorithms including SVM and random forest.

Keywords Yelp user review · Review analysis · Text mining · Sentimental analysis · Sentiment polarity classification · Star rating prediction · Machine learning · Naïve Bayes · Natural language processing (NLP)

1 Introduction

In this outgrowing usage of Internet, digital medium has picked up the highest reach in the sleeves of any individual in forms of social media Web sites. With the increasing user content on the Web, it becomes essential for the business heads to analyze the

A. Shah (✉) · K. Kothari · U. Thakkar · S. Khara
Silver Oak College of Engineering and Technology, Ahmedabad, India
e-mail: aagamshah96@gmail.com

K. Kothari
e-mail: komalkothari96@gmail.com

U. Thakkar
e-mail: umang.thakker21@gmail.com

S. Khara
e-mail: satvik@socet.edu.in

hidden potential reviews for a product in order to match the needs of the customer and to analyze the market growth for doing better improvement in the product. Data mining is a field which deals with big data to extract knowledge from Web and to store it efficiently. It has been widely observed that the feedback review provided on the app or Web site is not trustworthy as they do not reveal the real experience of the users. Many a times, it is seen that the product reviews are posted by spammers or AI bots in order to defame the particular organization by creating discrepancies in those reviews. Also, it has become vital to extract the text from all over the Web related to the product and then analyze it to find the real feedback to improve the product. This accurate analysis can be achieved by the concept of text mining.

Related Work: Several types of algorithms and methodologies are developed in the research done for sentimental analysis and machine learning. A study paper presented by Mouthami [1], says that, for regulating the task of classifying a single topic textual review, and for expressing a positive or negative evaluation, a new algorithm called Sentiment Fuzzy Classification, where parts of speech tags are used to improve the classification accuracy on the benchmark dataset of Movies reviews, is developed. While a study by Zeenia Singla [2] was based on statistical and sentimental analysis of consumer product reviews where user reviews were filtered on basis of two categories, positive and negative, on a large set of online reviews for mobile phones, and statistical analysis was conducted.

In this research paper, we are classifying user reviews in terms of multi-label evaluation using text mining and machine learning algorithms. Sentiment polarity classification will provide the true sentiments of the user reviews and the real meaning of the text written. While n-gram classification is done to avoid the misleading phrases in a sentence which may provide a different meaning. And machine learning algorithms will ensure us the predictive outcome from the given training dataset into a new test dataset. Natural language processing (NLP) word processing is done in order to evaluate the user reviews accurately.

We collected the dataset from the YELP, a leading user review provider company based in USA, which was openly accessible for academic purposes. The collected dataset, due to the randomness of the reviews' distribution according to all the businesses including hospitals, hotels, restaurants, cafes, and other business organizations, was then filtered using the concepts of text mining. Hereby, we focused our analysis on the user reviews for restaurants based in USA and Canada.

Below given is the process diagram of the applied methodologies for the project (Fig. 1).

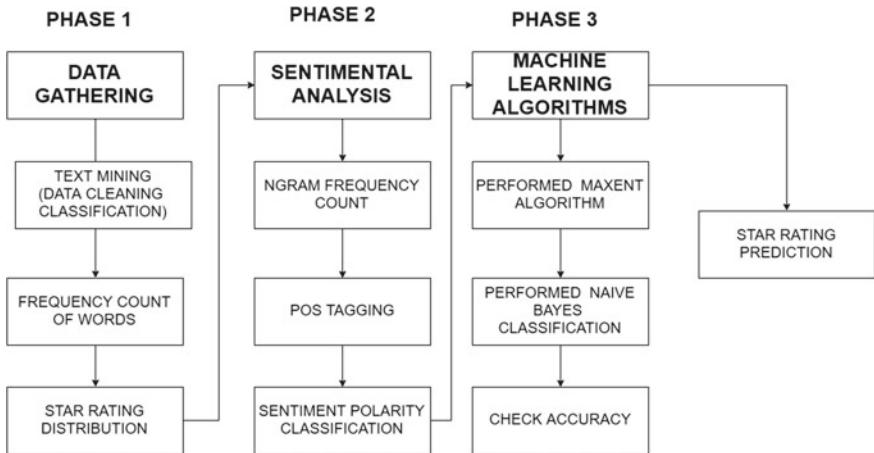


Fig. 1 Process diagram

2 Research Design and Used Methodologies (Proposed Approach)

2.1 Text Mining

Text mining is the use of high data processing methods and algorithms to obtain worthy knowledge from raw text. Text mining and analytics are the terms which portray a wide collection of methods for evaluating differently organized text data. The agenda behind each of these methodologies is the requirement to turn text into statistics, such that text mining algorithms can be implemented to huge databases, which require knowing of how to both use and combine techniques for handling text, ranging from individual words to documents to entire document databases [3].

Text mining for the user review analysis brought to us the need of having a dataset which possess data fields such as of User_name, Restaurant_name, Star_ratings, Overall_ratings, and User_review. We obtained csv files of multiple raw dataset from an online food application named YELP and started feature engineering with the help of text mining method: Data Cleaning. The huge volume of data was categorized and classified for proper evaluation of only restaurant reviews in American states. As a result, our final dataset consisted of content user reviews from different restaurants with detailed reviews and corresponding ratings.

Frequency count feature of text mining was performed in order to know the volume of the individual words used, and then the common English language words such as prepositions and conjunctions which were not beneficial in analysis were removed to improve accuracy. After that, most occurring words in a review were calculated for better judgement of analysis (Fig. 2).

Fig. 2 Frequency count for top words of Rating 4 (better)

word	better.freq
food	313
indian	268
good	245
buffet	213
chicken	197
naan	167
place	147
lunch	135
great	125

Table 1 User review distribution according to star ratings

Star ratings	Meaning
1 -	Worst
2 -	Bad
3 -	OK
4 -	Better
5 -	Best

Using *tm* package of text mining, word corpus was made after distributing the reviews into categorized star ratings as per Table 1. Star ratings ranging 1–5 were then split into different word corpuses according to its meaning.

These word corpuses are then divided according to the distribution of ratings which enhance as

Rating 1(Worst) and Rating 2(Bad) \Rightarrow **Negative**
 Rating 3(OK) \Rightarrow **Neutral**
 Rating 4(Better) and Rating 5(Best) \Rightarrow **Positive**

These bifurcations made us easy to count the exact word count of the particular rating and hence can further be helpful in prediction too.

2.2 Sentimental Analysis

Sentimental analysis can judge the sentiments of a person from the text content provided. Sentiment is a feeling, opinions, views, or perspectives driven by feelings. Sentimental analysis, which is otherwise called as opinion mining, acts upon individuals' notions toward specific entities. Internet is a resourceful mine regarding sentiment mining from the text [4]. Sentimental analysis focuses on understanding the motive of the user. The analyst gathers the data from social media, tweets, online forums, blogs, comments, and user reviews and then utilizes it in analyzing the real meaning and the opinion toward their product.

We have obtained the dataframe of positive-negative reviews for sentimental analysis. But before starting that, we had observed from other papers that sentimental classification cannot accurately analyze the meaning of the review if sarcastically written. For example,

A review like '*This place was so good that I would never come back here.*' will imply a positive sentiment even though the real meaning of the sentence possesses negative sentiments of the user. So in order to rectify that misclassification, we proposed n-gram classification and part-of-speech (POS) tagging on our YELP dataset of user reviews.

N-gram Classification:

N-grams are comprehensively used in natural language processing tasks. It is the aspect-level mining of sentimental analysis. They are technically a group of consequent words within a given set of text. While evaluating the n-grams, one word is moved forward making the last word of the previous n-gram as the first word for the next n-gram [5]. This kind of classification is beneficial in understanding the real aspect of the text. We have translated the positive-negative datasets into unigram, bigram, trigram, and quadgram for improving accuracy in the classification parameters (Fig. 3).

Part-of-Speech (POS) tagging:

The practice of distributing terms into their parts of speech and tagging them accordingly is known as POS tagging. Parts of speech are also known as word classes or lexical categories [6]. It is the aspect-level sentimental classification method. POS tagging evaluates each word from the sentence and then labels each word in form of whether it is noun, verb, conjunction, and etcetera. This helps in understanding the real meaning of the text and also be helpful in making a machine train. We used POS tagging for better and accurate prediction of user ratings after high-level text classification of each individual review from the raw dataset (Fig. 4).

Sentiment Polarity Classification:

After performing n-gram classification and POS tagging, the sentimental analysis is completed by obtaining sentiment score of each reviews. With the help of sentiment polarity classification, we obtained the polarity (positive/negative) of each word. The POS tagger gave us the benefit of handling the polarity classification by much improved accuracy and also provided sentiment score which was beneficial for training the dataset for machine learning purpose.

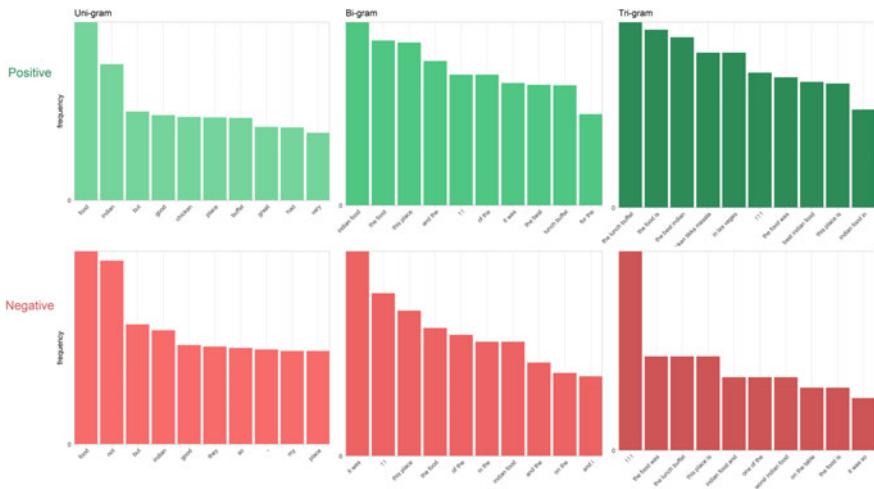


Fig. 3 Chart represents top 10 frequency count of n-gram classification for positive (green) and negative (red) words extracted from reviews

For example: "The ambience was good and I loved the food"

POStagged : The/DT ambience/NN was/VBD good/JJ and/CC
I/PRP loved/VBD the/DT food/NN

POStags: "DT" "NN" "VBD" "JJ" "CC" "PRP" "VBD" "DT" "NN"

TAG	DESCRIPTION
DT	DETERMINER
NN	NOUN,SINGULAR OR MASS
VBD	VERB, PAST TENSE
JJ	ADJECTIVE
CC	CORDINATING CONJUNCTION
PRP	PERSONAL PRONOUN

Fig. 4 POS tagging example and description

2.3 Machine Learning

Machine learning is a branch of artificial intelligence (AI) which is considered as the future of computing. It avails systems the capacity to automatically learn and progress from experience without being explicitly programmed [7]. Machine learning is training a model from a sample dataset and performing operations on it in order to make machine learn to do things by itself.

Machine learning algorithms are often categorized as supervised or unsupervised learning method. In our proposed approach, we have used supervised learning method

Fig. 5 Star rating prediction using MaxEnt model

true_rating	maxent_rating
5	5
5	4
5	5
5	5
5	4
5	4
5	4
4	4
5	5

for star rating prediction. We have combined two machine learning algorithms in order to perform prediction: maximum entropy model and Naïve Bayes classifier.

MaxEnt Model:

MaxEnt stands for maximum entropy model. Maximum entropy model is a text classification algorithm used for prediction purposes. It is based on the principle of maximum entropy. It has strength to learn and remember millions of features from sample data [8]. Maximum entropy methods are very general ways to predict probability distributions on given constraints. With the help of n-gram classification, that data is now taken as input to the MaxEnt model.

Due to the splitting of data into n-grams and the frequency calculations, MaxEnt model prediction algorithm was applied. MaxEnt model provides us the real ratings along with the original star ratings. This star rating prediction is nearly around with 90% precision when checked with the provided training and testing dataset's ratings (Fig. 5).

This shows the true rating, and the MaxEnt predicted rating after the proper analysis of the words obtained from n-gram classification method.

Naïve Bayes Classifier Model:

Naïve Bayes is one the most traditional methods used in both classification and prediction algorithms. Naive Bayes is a basic method for building classifiers: models that allot out class labels to problem categories, represented to as vectors of featured esteems from the raw data, where the class labels are drawn from some limited informational dataset. Naïve Bayes is not only for training such classifiers; it is a group of algorithms connected with a common Bayesian principle: All Naive Bayes classifiers accept that the estimation of a specific feature is independent of the corresponding value of some other feature, when given the class variable.

We have utilized the maximum of Naive Bayes prediction algorithm by applying it upon the dataset after performing sentiment polarity classification. It provided sentiment scores by calculating the frequency of positive and negative words in a user review from the positive-negative words dictionary; and then

```
A-priori probabilities:
Y
      1       2       3       4       5
0.06202532 0.03924051 0.06835443 0.20253165 0.62784810
```

Fig. 6 A priori probabilities of Ratings 1–5 in Naïve Bayes model

pred	1	2	3	4	5
1	64	0	0	2	11
2	0	39	1	1	7
3	0	0	76	0	7
4	0	0	0	201	22
5	2	0	3	42	674

Fig. 7 Matrix shows actual and predicted ratings combination

$$\text{SentiScores} = \text{Positive/Negative}$$

These sentiscores provide the real sentimental score of the individual review by properly evaluating the resultant dataframe obtained after POS tagging. This output was then used to predict the star ratings using Naïve Bayes model. Naïve Bayes model learnt the correspondence value of the sentiscores and the star ratings by matching them with the dataset and obtaining a priori probabilities for the same. A priori probabilities showed the probability of possibly having those ratings (Fig. 6).

This shows that these amounts of reviews show a tendency to possess the listed star ratings with Rating 5 having the highest number of reviews at 0.62784810 probability out of 1.

Furthermore, with the help of training as well as test dataset and the sentiscores provided of each individual review, Naïve Bayes prediction model was implemented. For implementing on the training dataset, multi-label classification was required to be taken as an input, which was also fulfilled by the Naïve Bayes model. The output provided the estimated ratings, from the machine learnt dataset of positive and negative words of each review (Fig. 7).

These show the actual and predicted ratings through Naïve Bayes model. As we can see, the diagonal values are increasing which proves that the actual and predicted ratings match to a greater extend, providing us the accuracy of 91.49%.

3 Result and Discussion

After combining two algorithms, and multiple classifiers, we have obtained the accuracy of 91.49% which was for the number of same occurrences between the predicted and the true ratings. After the entire research, we have pointed out our views that, due

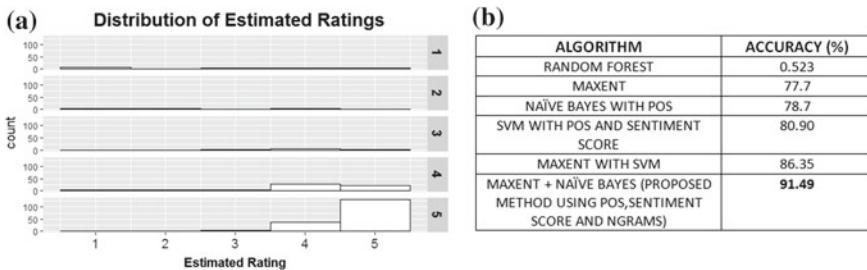


Fig. 8 **a** Graphical presentation of estimated ratings with the review count and **b** a comparison chart for the applied algorithms and obtained accuracies

to the imbalance in the nature of user reviews and the correspondent ratings, there is a conflict between prediction and actual ratings. Humans show a tendency to write something different review and give some different rating completely irrelevant to the written review. We propose our future work to solve the discrepancies between the remaining 8.6% of the true and predicted ratings in order to provide authenticity in the user reviews according to the corresponding ratings by working on a bigger and better dataset.

Below given are the graphical representation of the proposed methodology performed above and the comparison chart of the accuracy for all the algorithms applied with different approach on the same dataset (Fig. 8).

Conclusion: It is clearly observed that the combined methodology of performing MaxEnt and Naïve Bayes together with the help of n-gram classification and POS tagging provides estimation of ratings with highest accuracy.

References

1. Sentiment Analysis and Classification Based on Textual Reviews—IEEE Conference Publication. Ieeexplore.ieee.org. (2018). Retrieved 17 Feb 2018, from <http://ieeexplore.ieee.org/document/6508366/>
2. Statistical and Sentiment Analysis of Consumer Product Reviews—IEEE Conference Publication. Ieeexplore.ieee.org. (2018). Retrieved 19 Feb 2018, from <http://ieeexplore.ieee.org/document/8203960/>
3. Miner, G., Delen, D., Elder IV, J., Fast, A., Hill, T., Nisbet, R.: Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications [online]. Elderresearch.com. (2012). Available at: https://www.elderresearch.com/hubfs/Whitepaper_The_Seven_Practice_Areas_of_Text_Analytics_Chapter_2_Excerpt.pdf. Accessed 8 Jun 2018
4. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data* **2**(1) (2015). <https://doi.org/10.1186/s40537-015-0015-2>
5. What are N-Grams? [online]. Text-analytics101.rxnlp.com. (2019). Available at: <https://www.text-analytics101.rxnlp.com/2014/11/what-are-n-grams.html>. Accessed 25 Mar 2019
6. Categorizing and Tagging Words. Nltk.org. (2018). Retrieved 10 June 2018, from <https://www.nltk.org/book/ch05.html>

7. What is Machine Learning? A definition—Expert System. Expertsystem.com. (2017). Retrieved 9 June 2018, from <https://www.expertsystem.com/machine-learning-definition/>
8. Natural Language Processing—Maximum Entropy Modeling Algorithm and Sentiment Analysis Problem. Viblo. (2018). Retrieved 9 June 2018, from <https://viblo.asia/p/natural-language-processing-maximum-entropy-modeling-algorithm-and-sentiment-analysis-problem-xQMkJDjaGam>

Veracity Analysis and Prediction in Social Big Data



P. Suthanthira Devi, S. Karthika, P. Venugopal and R. Geetha

Abstract In big data sector, social media plays an important role for providing the biggest amount of unstructured data. Twitter is one such social media platform for dissipating unverified information during the time of disaster like cyclones, earthquake and flood. Veracity is the degree to which the information is accurate and trusted. The main objective of this paper is to identify the veracity of the rumour and find the source of the rumour. The authors analyse the PHEME data set in which the Sydneysiege event is chosen for analysis. This paper proposes a novel and hybrid rumour source–detector approach which combines the merits of graph and tree data structures. It identifies the source node traversing through general tree and predicts the rumour-spreading possibilities with general graph using spanning tree.

Keywords Big data · Veracity · Rumour source estimator · General graph · Rumour centrality

1 Introduction

Twitter is the most popular microblogging Web service. It has around 600 million active users. Nowadays, Twitter can be the most prominent news source to spread information to the mass community of people. It provides valuable information to the people in real-world emergency periods like floods and cyclone. However, this information contains noise and different forms of e-crimes like spreading rumours,

P. Suthanthira Devi (✉) · S. Karthika · P. Venugopal · R. Geetha
SSN College of Engineering, Chennai, India
e-mail: devisaran2004@gmail.com

S. Karthika
e-mail: skarthika@ssn.edu.in

P. Venugopal
e-mail: venugopalp@ssn.edu.in

R. Geetha
e-mail: geethkaajal@gmail.com

spam and phishing [1]. Rumour is defined as “circulating story of doubtful veracity, which is actually credible but hard to verify”. It crops plenty of optimism and anxiety to motivate finding out the reality [2]. Most research works demonstrate Twitter also spreads misinformation and fake news. For one rumour tweet, there are nearly eighty thousand active users who had been involved in both information sharing and diffusion. Takayasu et al. [3] examine the Japan earthquake event and show the real-time rumour diffusion and correction. In [4], the authors analyse Boston Marathon events, to identify the people who diffuse fake messages and characterize the Twitter attributes. During critical event time, it is important to detect the credible information from the Twitter [5].

The main contribution of this paper is (1) to identify the source of the rumour in the network and (2) to predict the possible chances of the rumour-spreading nodes in the network. In this paper, Sect. 2 discusses rumour centrality, fake news propagation and rumour source identification. Section 3 presents the graph theoretical approach using general tree for identifying the source of rumour. Section 4 designs graph with BFS structure to predict the spreading possibility of the rumour. Section 5 discusses the event called Sydneysiege and shows the effectiveness of the rumour source–detector. Section 6 concludes the proposed models.

2 Related Work

Rumour Centrality: Devavrat et al. [6] develop a graph scoring function rumour centrality, used to rank the nodes as a source. For regular trees, rumour centrality is an ML estimator, and the SI models are very effective for calculating and spreading time distributions and the heterogeneity in the tree structure. The estimated source probability is higher than k-hops distant from the original source decays exponentially in k.

Rumour Source Identification: Tauhid et al. [7] identify the source of the tweet and retweet using prediction features. They present a system for predicting sole retweets in Twitter. Twitter data are trained by the probabilistic collaborative filtering models to predict the tweets and retweets. In [8], the authors are detecting the source of the rumour in a network based on an uninfected and infected nodes (SIR Model). The authors of [9, 10] are constructing the probabilistic model and derivation for rumour source estimator. They provide a derivation for rumour source estimator as a regular tree, general trees and graph. The authors do not work on the real data set; to overcome this problem, the author used real-time data set for one critical event called Sydneysiege used to identify the source of the rumour and further who is going to spread the rumour.

3 Rumour Source–Detector: Moving from Regular Tree to General Tree

Definition 1 Let $G(S, E)$ be a graph constructed based on social network, where S is the number of the nodes in the network and E is an edge of the form (p, q) for some p and q in S . Suppose a rumour started from the source node say s^* at a particular time t and circulated in the network G , then the rumour source–detector \hat{s} is defined as

$$\hat{s} \in \arg \max_{s \in G_N} P(G_N/s) \quad (1)$$

where ‘ n ’ is the number of the infected nodes at some time t . Assume node s^* is infected at time t . G_N represents the connected graph with n infected nodes. \hat{s} is the detector for the infected node. $P(G_N/s)$ represents the probability of G_N getting of the ‘ n ’ nodes that are infected, starting ‘ s ’ as the source node.

ML Detector for Regular Tree

STEP 1: Find the set of all permitted permutations $\Omega(s, G_N)$ that form the rumour graph G_N starting with the source node S .

STEP 2: For each $\sigma \in \Omega(s, G_N)$, the probability of $P(\sigma/s)$ is given by

$$P(\sigma/s) = \prod_{k=2}^N P(K^{\text{th}} \text{ infected node} = s_k / G_{k-1}(\sigma)/S) \quad (2)$$

STEP 3: If the graph G_{k-1} has $n_{k-1}(\sigma)$ uninfected nodes, then the possibility of uninfected node to be infected is

$$\frac{1}{n_{k-1}(\sigma)} \quad (3)$$

STEP 4: Probability of all $k - 1$ nodes to be infected is given by

$$P(\sigma/v) = \prod_{k=2}^N \left(\frac{1}{n_{k-1}(\sigma)} \right) \quad (4)$$

$$P(\sigma/s) = \prod_{k=2}^N \frac{1}{d_1(\sigma) + \sum_{i=2}^k (d_i(\sigma) - 2)} \quad (5)$$

STEP 5: In regular tree, all the nodes have the same degree ‘ d ’ and hence all permitted permutations σ have equal probability, independent of the source. In this case, Eq. (5) becomes the probability of

$$P(\sigma/v) = \prod_{k=1}^{N-1} \frac{1}{dk - 2(k-1)} \equiv p(d, N) \quad (6)$$

In real-time applications, general tree plays a major role than regular trees [11]. In general tree, every node has different probabilities so that the rumour spreads in different permitted permutations. General tree first fixes the source node and then arranges all the nodes based on BFS algorithm. If σ_s^* is the permitted permutation starting from the source node s , then the rumour source-detector is given by

$$\hat{s} = \arg \max_{s \in G_N} P(\sigma_s^*|s) \cdot R(s, G_N) \quad (7)$$

where $P(\sigma_s^*|s)$ is the probability of permitted permutation and $R(s, G_N)$ is the total number of the distinct permitted permutation of nodes G .

4 Rumour Source Detector: General Graph

Definition 2 Let $G(s, E)$ be the general graph for $s \in G_N$. Assume s^* is the source node for rumour. This rumour spreads with breadth-first search tree rooted at s , $T_{\text{bfs}}(v)$, and the rumour source-detector for a general graph is given by

$$\hat{s} = \arg \max_{s \in G_N} P(\sigma_s^*|s) R(s, T_{\text{bfs}}(s)) \quad (8)$$

$P(\sigma_s^*|s)$ represents the probability of permitted permutation, and $R(s, T_{\text{bfs}}(s))$ represents the rumour spreads with breadth-first search (BFS) tree rooted at $(s, T_{\text{bfs}}(s))$ where $T_{\text{BFS}}(s)$ is breadth-first search tree of node S .

5 Results and Discussion

The PHEME data set contains Twitter conversations which are initiated by a rumourous tweet, and the conversations include tweets responding to those rumourous tweets [12]. In this paper, the authors analyse a particular Sydneysiege event to identify the source of the event #Sydneysiege.

Tweet: *The gunman hostage in December 2014, where 10 customers and 8 employees of chocolate café at martin place in Sydney.*

The annotation and the preprocess of Sydneysiege event is performed using the following steps

Step 1: Large data set was sampled automatically.

Step 2: All conversational threads like retweets and direct and indirect reply were collected.

Step 3: Manual classification is used to segregate the rumour and non-rumour data.

SydneySiege event classification is explained in Fig. 1.

SydneySiege event contains 71 conversational threads. The conversational thread features like misinformation, isturnaroundtime and true are used to classify the rumours [13]. There are 7 rumour threads and 64 non-rumour threads. Further classification done using support type, in this 7 rumour threads, 4 threads are support the rumour and 3 threads are not support. Manual annotation of the SydneySiege, the TweetID 9121, is the source of the rumour. The proposed hybrid rumour source-detector also identifies the same TwitterID as the source.

If source tweet contains rumourous story, then the followers can reply to the parent tweet, in favour (or) against the source tweet. The response type has four categories, Agreed—repliers support the source information, Disagreed—deny the source tweet, Appeal for more information and Comment—reply own comment. The two types of responses are direct, and deep reply is shown in Table 1.

Rumour Source-Detector for Source Identification:

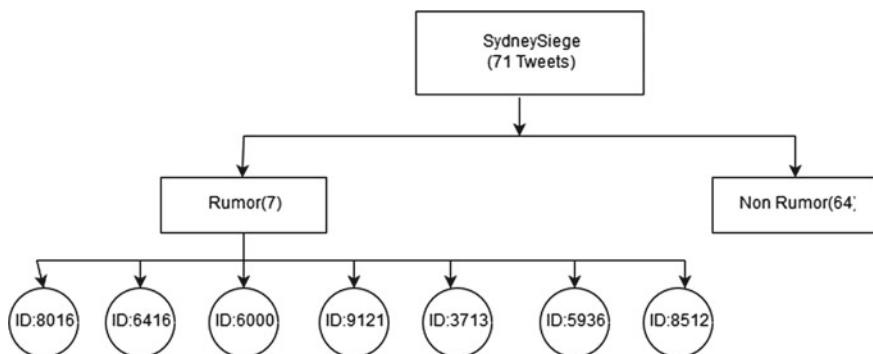


Fig. 1 SydneySiege event conversational threads

Table 1 Classification of SydneySiege event threads

Type	Agree	Disagree	Comment	Appeal for more information
Direct reply (response type vs. source)	0	6	6	Nil
Deep reply (response type vs. previous)	1	9	12	2
Deep reply (response type vs. source)	5	3	12	4

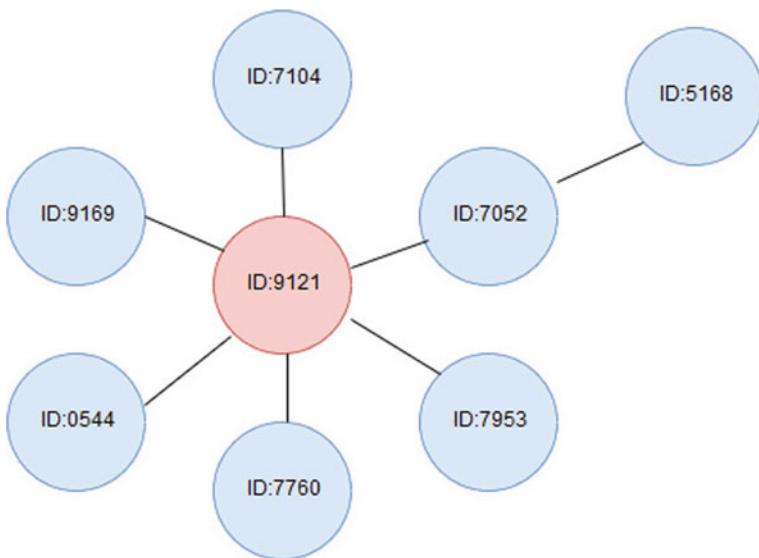


Fig. 2 Tree structure for rumour nodes

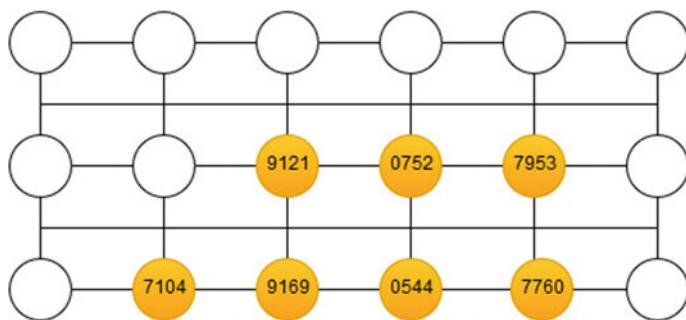


Fig. 3 Rumour network-infected nodes are shaded

The authors identified TwitterID 9121 has the highest number of reply from the supporting threads and marked it as the source of the rumour. Figure 2 shows the tree structure of rumour nodes for Sydneysiege event.

Rumour source-detector in general tree is calculated for Sydneysiege event and is shown in Table 2.

Therefore, $P(c_{9121}^*)$ is simply two times likely as nodes 2, 3, 4, 5, 6 to be the source. The probability calculation shows that the node 9121 is the source node than other 6 nodes. So, the TwitterID 9121 is concluded to be the rumour source node.

Rumour Source-Detector for Predicting the Spread of Rumour:

Figure 3 shows the rumour network with infected nodes for Sydneysiege event.

The nodes are arranged in BFS tree structure with spanning tree.

Table 2 Rumour source detection using tree

For node 9121	For node 0752	For node 7953,7760,0544,9169
$\hat{s} = \arg \max_{s \in Gn} P(\sigma_s^* s) R(s, Gn)$ $R(\sigma_{9121}^*) = \left(\frac{1}{6}\right)^6$ $R(s, Gn) = 6!$ $\hat{s} = \left(\frac{1}{6}\right)^6 * 6!$ $= \left(\frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6}\right) * 720$ $= 20\left(\frac{1}{6}\right)^4$	$P(\sigma_{0752}^*) = \frac{1}{2}\left(\frac{1}{6}\right)^5$ $R(v, Gn) = 5!$ $\hat{V} = \frac{1}{2}\left(\frac{1}{6}\right)^5 * 5!$ $= \frac{1}{2} * \left(\frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6}\right) * 120$ $= 10\left(\frac{1}{6}\right)^4$ $= 20\left(\frac{1}{6}\right)^3$	$P(\sigma_{7953}^*) = 1 * \left(\frac{1}{6}\right)^5$ $R(v, Gn) = 5!$ $\hat{V} = 1 * \left(\frac{1}{6}\right)^5 * 5!$ $= 1 * \left(\frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6} * \frac{1}{6}\right) * 120$

BFS : 9121, 0752, 7953, 7760, 0544, 9169, 7104

In this network, every node has degree $d = 4$, and then the probability of the permitted permutation is calculated.

$$P(\sigma_1^*/1), P(\sigma_2^*/1)P(\sigma_3^*/1)P(\sigma_4^*/1) = \prod_{k=2}^{n=4} \frac{1}{2(k+1)} = \frac{1}{4 * 6 * 8 * 10}$$

From Eq. (8), the \hat{s} is computed by

$$\hat{s} = P(\sigma_1^*|1)R(s, T_{\text{bfs}}(s)) = \frac{1}{4 * 6 * 8 * 10} * \frac{7!}{7 * 1 * 1 * 1 * 2 * 2 * 2}$$

Similarly, rumour source-detector can be calculated for other nodes. Figure 4 shows the BFS tree for all rumour nodes in the network. Estimating all probabilities, the nodes (9169, 0544) maximize the probability and act as the rumour source for the given network. The authors observed that predicting the possibility of the rumour-spreading nodes in the network is highly efficient using graph theory techniques.

6 Conclusion

The authors design a hybrid rumour source-detector by using both general tree and general graph. This paper analyses various features like misinformation, true, isturnarountime of the Twitter conversation threads, to classify the rumour and non-rumour for one specific event Sydneyseige. The proposed hybrid rumour source-detector \hat{s} accurately identifies the TwitterID 9121 as the source node which matches the manual evaluation performed on the data set. Similarly, graph rumour source-detector has predicted the TweetID (9169, 0544) as the possible rumour-spreading nodes and this also matches the manual evaluation. The authors conclude that the graph theory approach is highly effective in veracity analysis of social big data.

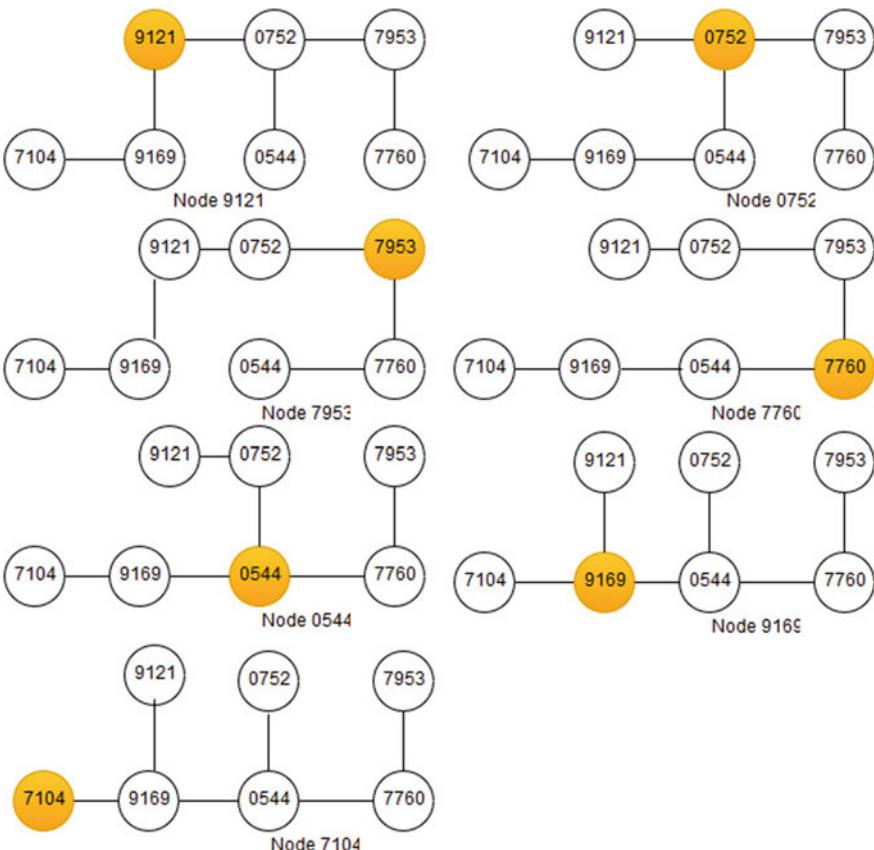


Fig. 4 BFS tree for each node

References

1. Symantec Corporation. Istr: Internet Security Threat Report 2013
2. Zubiaga, A., Liakata, M., Procter, R., Bontcheva, K., Tolmie, P.: Towards detecting rumours in social media. In: AAAI Workshop: AI for Cities (April 2015)
3. Takayasu, M., Sato, K., Sano, Y., Yamada, K., Miura, W., Takayasu, H.: Rumour diffusion and convergence during the 3.11 earthquake: a Twitter case study. *PLoS one* **10**(4), e0121443 (2015)
4. Gupta, A., Lamba, H., Kumaraguru, P.: \$1.00 per rt #bostonmarathon #prayforboston: analyzing fake content on Twitter. In: Proceedings of the Eighth IEEE APWG eCrime Research Summit (eCRS), p. 12. IEEE (2013)
5. Gupta, A., et al.: Tweetcred: real-time credibility assessment of content on Twitter. In: International Conference on Social Informatics. Springer, Cham (2014)
6. Shah, D., Zaman, T.: Rumour centrality: a universal source detector. In: ACM SIGMETRICS Performance Evaluation Review, vol. 40, no. 1. ACM (2012)

7. Zaman, T.R., Herbrich, R., Van Gael, J., Stern, D.: Predicting information spreading in Twitter. In: Workshop on computational social science and the wisdom of crowds, nips, vol. 104, no. 45, pp. 17599–17601. Citeseer (December 2010)
8. Qazvinian, V., Rosengren, E., Radev, D.R., Mei, Q.: Rumour has it: identifying misinformation in micro blogs. In: Proceedings of EMNLP, p. 1589–1599 (2011)
9. Procter, R., Vis, F., Voss, A.: Reading the riots on Twitter: methodological innovation for the analysis of big data. *Int. J. Soc. Res. Methodol.* **16**(3), 197–214 (2013). <https://doi.org/10.1080/13645579.2013.77417223>
10. Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R.M.: Rumours, FalseFlags, and digital vigilantes: misinformation on Twitter after the 2013 Boston Marathon Bombing. In: Proceedings of iConference.iSchools (2014)
11. Shah, D., Zaman, T.: Rumors in a network: who's the culprit? *IEEE Trans. Inf. Theory* **57**(8), 5163–5181 (2011)
12. Lukasik, M., Cohn, T., Bontcheva, K.: Classifying tweet level judgements of rumours in social media. arXiv preprint [arXiv:1506.00468](https://arxiv.org/abs/1506.00468). (2015)
13. Zubriaga, A., Liakata, M., Procter, R., Hoi, G.W.S., Tolmie, P.: Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE* **11**(3), e0150989 (2015)

Efficient Implementation of Carry-Skip Adder Using CSMT Adder and PPA



S. Sridevi Sathya Priya, Benita Poulraju, B. Benita
and Dharani Karanatakam

Abstract In this paper, carry-skip adder is implemented using carry-select modified-tree (CSMT) adder and parallel prefix adder (PPA) topologies. This implementation is suitable for low-power application. The CSMT adder is designed using multiplexers. This adder uses very few numbers of multiplexers in their design, thereby low power can be achieved. PPA structure decreases number of stages in the design as well as smaller delays. The results of the implementation are compared with the conventional method in terms of power, delay and number of gates.

Keywords Carry-skip adder · CSMT adder · Low-power adder · Efficient carry-skip adder

1 Introduction

The arithmetic and logical blocks are used to perform many logical and arithmetic operations such as addition, multiplication, division and subtraction. Among all the operations, the binary addition is considered as a basic operation. Using this operation, many other operations can be done. The binary parallel adder adds two binary numbers using copies of binary full adders with a specified speed. The generation of the carry and its propagation delay in a binary adder are the main problems in such adders. The schemes which are used in fast carry generation [1] during addition process can be used to design high-speed binary adder. These high-speed adders can be implemented using multiplexers [2]. Thereby, the overall arithmetic operation's processing speed can be increased because for most of the arithmetic operation, addition is the basic operation. Fast binary adders can also be designed to determine the similarity between redundant to binary conversion (RBC) and binary addition operation. The redundant digits which are the bits of adder operands must be converted into a binary number using RBC. Multiplexers are used to construct RBC adder by

S. Sridevi Sathya Priya (✉) · B. Poulraju · B. Benita · D. Karanatakam
ECE, Karunya Institute of Technology and Sciences, Coimbatore, India
e-mail: s.d.s.priya@gmail.com

proper recording of the input bits. These types of adder use two types of fast carry generators, known as tree-based and carry-select. The carry-select adders with many blocks of smaller block lengths are designed for specified latency. This consumes less power. There were efforts taken to design low-power adders with least energy consumption and area consumption.

2 Related Work

Initially, the parallel adders were designed and the performances were analysed [3]. These adders are designed using uniform static CMOS layout methodology which is used to reduce the short-circuit power. Chirca et al. designed carry-skip adder with low power and high performance [4]. Low power is achieved by reducing the number of logic levels, glitches and transistors in the design and performance is increased by using optimized carry look-ahead adders. To achieve further low power, a hybrid carry-select modified-tree (CSMT) adder architectures is used for binary carry generators and adders are implemented using multiplexers only. To further decrease the delay, addition operation can be achieved by parallel prefix adders (PPA). These adders are used in implementing fast adders [6]. In this paper, to increase the speed, PPA is used in implementation and to reduce the power, CSMT adder with multiplexer implementation is used.

3 Proposed Methodology

A ripple-carry adder delay can be reduced by using carry-skip adder. A carry-skip adder structure is designed using CSMT adder and PPA. This proposed architecture is designed to reduce the power consumption of the binary addition and also to improve the speed with less area. PPA is used in this implementation for smaller area and for achieving less power consumption. CSMT architecture is designed using the fewest number of multiplexers so it consumes less energy for a specified latency. The equations used for CSMT design is shown in Eq. (1).

$$\begin{aligned} C_{i+1} &= C_i b_{i'r} + \bar{C}_i a_{i'r} \\ S_{i+1} &= C_i (\overline{b_{i,r}} + a_{i'r}) \\ S_{i+1} &= C_i (\overline{b_{i,r}} + a_{i'r}) \end{aligned} \quad (1)$$

PPA is different from other adders due to its carry processing technique. In PPA, the carry is obtained in parallel so this leads to fast addition arithmetic operation. There are different types of PPA available; each type is exclusively different from each other depending on their carry graph and number of levels. The different types of PPA are namely, FCO, Brent–Kung adder and Kogge–Stone adder [10]. The huge

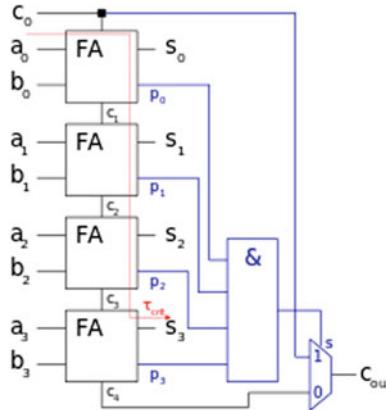


Fig. 1 Carry-skip adder block

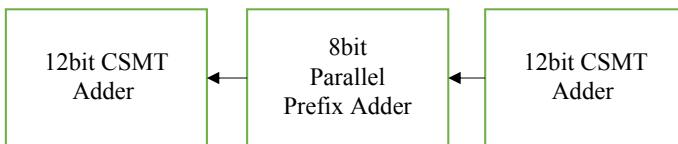


Fig. 2 Carry-skip adder using PPA and CSMT adder

number of levels are used in Brent–Kung adder (BKA). This reduces the overall speed of the adder. This is also power efficient because it occupies less area. Also this uses forward path for calculating longest carry. So, the longest carry can be calculated fast. The intermediate carries calculated backward path. Figure 1 shows the carry-skip adder. This carry-skip adder is considered for implementation and this is implemented using CSMT and PPA.

A 32-bit carry-skip adder is designed using 24-bit CSMT, and 8-bit parallel prefix adder is shown in Fig. 2.

4 Results and Discussions

A low-power carry-skip adder is designed and implemented using cadence tool under 180 nm technology. This implementation is done by using PPA and CSMT adder and the results are analysed. Figure 3 shows the schematic of the carry-skip adder using CSMT and parallel prefix adder. Figure 4 shows the simulation result of the carry-skip adder.

From Table 1, it can be concluded that proposed adder is more efficient as it consumes less power and delay when compared to conventional 32-bit carry-skip

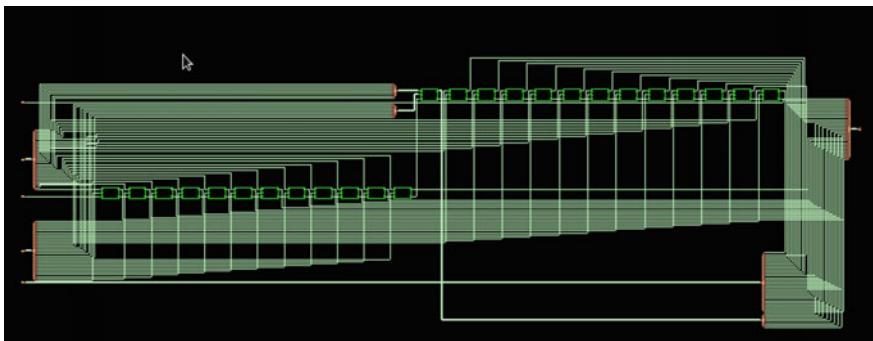


Fig. 3 Carry-skip adder using CSMT and parallel prefix adder

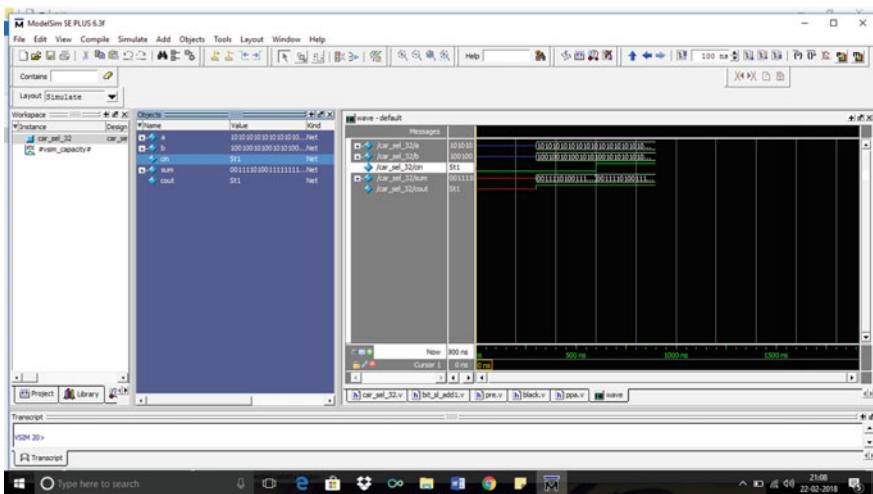


Fig. 4 Carry-skip adder using CSMT and parallel prefix adder

Table 1 Power delay and area comparison

Design	Power	Delay	Gates
32-bit carry-skip adder	397346.335 nW	2860 pS	6120
Proposed adder	389461.663 nW	721 pS	2390

adder. The power reduction is achieved by using CSMT and PPA is used to increase the speed.

5 Conclusion

This paper deals with the analysis of various low-power adder methodologies and their performances are compared. A 32-bit carry-skip adder is designed using 24-bit CSMT and 8-bit parallel prefix adder. The results are compared with CSMT adder and PPA provides better performance than the conventional adder. Using this proposed methodology, 389461.663 nW power is achieved with 74% improvement in speed and with 60% reduction in area.

References

1. Ramkumar, B., Kittur, H.M.: Low-power and area-efficient carry select adder. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **20**(2) (Feb 2012)
2. Kumar, A., Singh, A.S.: Full adder using multiplexer. *IJIRT* **1**(12) (2015)
3. Nagendra, A., Irwin, M.J., Owens, R.M.: Area-time-power tradeoffs in parallel adders. *IEEE Trans. Circuits Syst. II* **43**, 689–702 (1996)
4. Chircă, K., Schulte, M., Glossner, J.: A static low-power, high-performance 32-bit carry skip adder. In: *IEEE Conference on Digital System Design* (2004)
5. Parhi, K.K., Low-energy CSMT carry generators and binary adders. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **7**(4) (Dec 1999)
6. Chaitanyakumari, P., Nagendra, R.: Design of 32 bit parallel prefix adders. *IOSR J. Electron. Commun. Eng.* **6**(1), 1–6 (May–Jun 2013)
7. Gizopoulos, D., Psarakis, M., Paschalidis, A., Zorian, Y.: Easily testable cellular carry look ahead adders. *J. Electron. Test. Theor. Appl.* **19**, 285–298 (2003)
8. Matsunaga, T., Kimura, S., Matsunaga, Y.: Power-conscious syntheses of parallel prefix adders under bitwise timing constraints. In: *Proceedings of the Workshop on Synthesis and System Integration of Mixed Information technologies (SASIMI)*, pp. 7–14. Sapporo, Japan (October 2007)
9. Choi, Y.: Parallel prefix adder design. In: *Proceedings of the 17th IEEE Symposium on Computer Arithmetic*, pp 90–98, 27th June 2005
10. Zamhari, N., Voon, P., Kipli, K., Chin, K.L., Husin, M.H.: Comparison of parallel prefix adder (PPA). In: *Proceedings of the World Congress on Engineering 2012*, vol. II WCE 2012, London, U.K., 4–6 July 2012

Multispectral Satellite Image Classification Based on Bare Bone Fireworks Algorithm



Eva Tuba, Raka Jovanovic and Milan Tuba

Abstract Satellite image classification is an important part of applications in various fields such as agriculture, environmental monitoring, and disaster management. K-means algorithm is a simple clustering method that can be adjusted for classification. Due to the fact that k-means represents a local search around the initially generated solutions, it should be combined with some global search method. We propose recent bare bone fireworks algorithm for k-means optimization used for image classification. The proposed method was tested on standard benchmark datasets and compared the results with other methods from the literature. Simulation results showed that the proposed combined approach is better for image classification compared to the original k-means algorithm, three other classification algorithms, and three methods based on other nature-inspired algorithms.

Keywords Multispectral satellite images · Optimization · Metaheuristics · Bare bone fireworks algorithm · Swarm intelligence · Classification

This work was supported by the Ministry of Education, Science and Technological Development of Republic of Serbia, Grant No. III-44006.

E. Tuba · M. Tuba (✉)
Faculty of Informatics and Computing, Singidunum University, Belgrade, Serbia
e-mail: tuba@ieee.org

E. Tuba
e-mail: etuba@ieee.org

R. Jovanovic
Qatar Environment and Energy Research Institute Hamad Bin Khalifa University,
Doha, Qatar
e-mail: rjovanovic@hbku.edu.qa

1 Introduction

In recent years, satellite images and their usage became one of the frequent research topics. Satellite images are used in various scientific fields and applications such as agriculture, geology, disaster management, military, and many more. They represent an important part of research and development in these fields.

Nowadays, most of the satellites capture multispectral images. Multispectral image represents one scene captured by different sensors where each recognizes specific wavelength range (band) including visible light and wavelengths around it. These bands are used for different purposes. For example, from 450 to 515 nm is the blue band that can be used for atmosphere and water analysis, green band from 515 to 590 nm is useful for investigating vegetation, red band from 590 to 690 nm, near-, mid-, and far-infrared bands are used for soil analysis, forest fires detection, geological features, etc. [1].

Since satellite images are part of numerous applications, different problems occur and most of them represent active research topics. One of the common tasks with satellite images is their classification. Depending on the purpose of images, classification can consider different types of crops, determining the growth stage of crops, different soil types, urban or mountain areas, etc.

In this paper, we addressed a problem of soil classification from satellite images. Soil type detection has great importance in agricultural and other technological uses of soils [2].

In order to classify multispectral satellite images, many different methods were proposed in the past. In [3], semi-supervised fuzzy k-means clustering algorithm was proposed for satellite image classification and change detection. In [4], neural network-based methods for remote sensing data analysis were proposed where unsupervised learning algorithm was used for feature selection. Another approach based on neural network is proposed in [5] where pixelwise classification of satellite images with convolutional neural networks was presented.

This paper presents a method for image classification based on k-means algorithm combined with a bare bone fireworks algorithm (BBFWA). K-means represents an unsupervised learning algorithm used for data clustering, while classification is a supervised learning problem. For supervised learning problems, the data labels are known and the goal is to make a model based on that data in order to assign labels to new instances. On contrary to supervised learning, unsupervised learning algorithms create models that assign labels to unknown data based on some similarity measure without using any knowledge about data labels or classes. Even though the differences are as presented, k-means algorithm can be adjusted and used for classification problem [6]. Each cluster represents one class. In k-means algorithm, instances are assigned to clusters based on similarities, usually distance, between them, and cluster centers. In the case of classification, similarity should be affected by distance but also should include instance labels.

Since k-means method has a significant drawback of getting trapped into local optima instead of the global optimum, it is usually combined with some other meth-

ods. In [7], k-means algorithm was combined with neural network for satellite image classification. In recent years, k-means algorithm was frequently combined with nature-inspired algorithms such as brain storm optimization algorithm [8, 9], firefly algorithm [10–12], and elephant herding optimization [13, 14]. Combined algorithms were also applied to image classification, e.g., artificial bee colony [15], bat algorithm [16], brain storm optimization [17], firefly algorithm [18], etc.

The proposed method was tested on standard benchmark datasets from the UCI repository [19] and compared to the original k-means, three nature-inspired classification methods from [16], and three classification algorithms from [20].

This paper consists of the following sections. Section 2 describes bare bone fireworks algorithm, while our proposed combined algorithm is explained in Sect. 3. Simulation results are presented in Sect. 4. The paper is concluded in Sect. 5.

2 Bare Bone Fireworks Algorithm

Bare bone fireworks algorithm (BBFWA) is a version of the firework algorithm presented by Li and Tan in 2018 [21]. It is a simplified version which reduced complexity but still remained efficient as proven in [21]. FWA and its modifications were applied to different real-world problems such as SVM parameters tuning in [22], image registration [23], and image compression [24].

BBFWA uses one firework to save the best solution from one generation to another. Search space bounded by the lower and upper bounds, i.e., d-dimensional vectors L_b and U_b , has been explored by the constant number of sparks that are produced by the firework's explosion. The best solution is represented as the position of the firework which is a d-dimensional vector x . In BBFWA, search space is analyzed by generating n random solutions ($s_i, i = 1; 2; \dots; n$) that are placed in the neighborhood of the best solution. The neighborhood is defined by the d-dimensional rectangle bounded by $x - A$ and $x + A$, where A represents an algorithm's parameter. This parameter A is used as a mechanism to perform exploration and exploitation. Size of the neighborhood is increased or reduced at the beginning of each generation by factor $C_a > 1$ or $C_r < 1$ depending on the best result found. If the best solution was improved in one generation, it can be assumed that the search space needs to be explored more, and thus, parameter A and the neighborhood will be increased by a factor of C_a . Opposite to that, in case that the best solution remains the same, it can indicate that the good area is found and that it needs to be analyzed better. In that case, the parameter A (and by that the neighborhood, too) is reduced by a factor C_r .

3 The Proposed Combined Classification Algorithm

Our proposed method for multispectral satellite image classification is a combination of the k-means and the bare bone fireworks algorithm. Quality of the results obtained

by the k-means algorithm is defined by the initial position of centroids, and this is not a good characteristic of the classification algorithm. The k-means algorithm can be improved by starting it several times with different initial solution and using the best one as the final solution. We propose guiding k-means algorithm by the BBFWA rather than using previously described random search.

Image classification can be defined as an optimization problem in the following way. Search space is limited by the instances. The optimization method searches for k centroids that will represent classes. Each centroid is described as d -dimensional vector, i.e., the number of attributes that describe instances. Since k centroids are searched, dimension of the optimization problem is $k * d$. The objective of the k-means algorithm used for clustering is to minimize the sum of distances between instances and their corresponding centroids. Since classification is a supervised learning problem, we need to incorporate the knowledge about the labels. The goal is to create model where instances of one class are in one cluster. If we consider one class, the representer (center) of that class can be arithmetical mean of instances in that class. Problem is that center should also distinguish instances from other classes, so in some cases, center should be further from instances so the instances from other class are not wrongly recognized. In order to do that, the following objective function is defined [15, 16, 18]:

$$\text{fit} = \frac{1}{D} \sum_{j=1}^D d(x_j, p_i^{c_i}), \quad (1)$$

where D is the number of instances used for training, and $p_i^{c_i}$ represents centroid of the class where x_j belongs based on the information from training set.

Solutions that represent coordinates of the centroids are generated and searched by BBFWA which will ensure global search. Pseudo-code of the proposed method is described in Algorithm 1.

Algorithm 1 Our proposed algorithm for multi-spectral satellite image classification

Require: web intelligence data, number of clusters k , parameter of the BBFWA algorithm C_a and C_r , number of sparks n , search space boundaries Lb and Ub

```

Generate random solution  $x$  from  $U(Lb, Ub)$ 
Evaluate generated solutions by Eq. (1)
Set the neighborhood  $A = Ub - Lb$ 
repeat
  for  $i = 1$  to  $n$  do
    Generate random solution  $s_i$  in range  $U(x - A, x + A)$ 
    Generate solution  $s_i$  (by mapping operator)
    Evaluate generated solution by Eq. (1)
  end for
  if  $\min_{i=1,2,\dots,n} (f(s_i)) < f(x)$  then
     $x = \text{argmin}(f(s_i))$ 
     $A = C_a A$ 
  else
     $A = C_r A$ 
  end if
until stopping criteria

```

4 Simulation Results

The code for the proposed BBFWA algorithm for multispectral satellite image classification was written in MATLAB version R2016b. All tests were run on Intel® Core™ i7-3770K CPU at 4 GHz, 8 GB RAM computer with Windows 10 Professional OS.

The proposed classification method was applied to the standard benchmark data for multispectral satellite image classification, Statlog (Landsat satellite that is freely available at the UCI Machine Learning Repository [19]).

The Landsat satellite dataset was obtained from NASA data used by the Australian Centre for remote sensing. It was created by taking sections of size 82×100 from the images captured by Landsat multispectral scanner. Features (36 features) are extracted from four different bands: two visible, green and red, and two from infrared bands. Extracted features are used to differentiate six classes of soil. The proposed method was trained and tested on different instances. Details about the dataset are presented in Table 1.

Classification quality was tested in the same way as in [16] where classification error percentage (CEP) was used which represents the percent of misclassified samples.

Another measure of the quality that was used in [16] is classification efficiency. A matrix of size $n \times n$, where n is the number of classes, is obtained by the classification algorithm. In that matrix, an entry q_{ij} stands for the number of instances that belong to class i are classified as class j . A perfect classification is performed if the classification

Table 1 Information about Landsat satellite dataset

Class no.	Class name	Training pixels	Test pixels
C1	Red soil	1072	461
C2	Cotton crop	479	224
C3	Gray soil	961	397
C4	Damp gray soil	415	211
C5	Soil with vegetation stubble	470	237
C6	Very damp gray soil	1038	470
	Total	4435	2000

Table 2 Comparison of the results for Landsat dataset

Measure	MLP BPNN	SVM	k-NN	BBFWA
CEP	10.7	14.9	11.0	8.7
μ_1	97.6	99.3	98.0	99.2
μ_2	97.3	93.3	95.5	98.2
μ_3	93.1	94.0	88.2	95
μ_4	65.9	42.2	66.8	67.8
μ_5	84.0	75.5	88.6	85.3
μ_6	88.3	83.8	87.9	88.8
μ_a	87.7	81.4	87.5	89.1
μ_o	89.3	85.1	89.0	91.3

matrix is diagonal. The individual (μ_i), average (μ_a), and overall (μ_o) efficiency of class i can be determined by the following equations:

$$\mu_i = 100 \frac{q_{ii}}{\sum_{j=1}^n q_{ji}}, \quad \mu_o = 100 \frac{1}{N} \sum_{i=1}^{n_c} q_{ii}, \quad \mu_a = \frac{1}{n_c} \sum_{i=1}^{n_c} \mu_i. \quad (2)$$

In order to test the quality of the proposed method, we compared our results with the methods proposed in [20] where three different classifiers were proposed for multispectral satellite image classification: multilayer perceptron backpropagation neural network (MLP BPNN), support vector machine (SVM), and k-nearest neighbor (k-NN). In [20], authors reported different quality metrics, but based on these results we could calculate measures used in this paper and presented in Table 2. The best results are printed in bold.

Classification error percentage obtained by our proposed method is 8.8, while the second-best classifier was neural network proposed in [20]. The proposed method was outperformed only for fifth class where k-NN obtained better individual efficiency.

Table 3 Comparison of the results for image segmentation dataset

Measure	K-means	GA	PSO	BA	BBFWA
CEP	41.38	32.68	32.45	25.90	24.04
μ_1	56	73	42	80	80
μ_2	100	99	98	100	100
μ_3	59	6	46	20	43
μ_4	0	56	61	77	78
μ_5	55	63	67	71	72
μ_6	86	87	84	82	83
μ_7	55	82	78	82	85
μ_a	58.7	66.6	68	74.7	77.3
μ_o	58.6	67.3	67.6	74.1	75.9

Based on the results presented in Table 2, we can conclude that our proposed method outperformed all three classifiers proposed in [20].

Besides classification of soils based on satellite images, often finer details need to be recognized such as paths, walls, and grass. Because of that, in [16] authors additionally tested their method on image segmentation dataset, also from the UCI repository [19]. We also included this dataset in our experiments. Image segmentation dataset contains 210 instances for training and 2100 test examples. Images should be classified into seven classes: brickface, sky, foliage, cement, window, path, and grass. There are 30 examples of each class for training the model, while 300 instances of each class are used for testing.

Results generated by applying our proposed method to the dataset were compared with the results of the method used in [16] where the proposed bat algorithm (BA) for classification was compared to the three other methods, genetic algorithm (GA), particle swarm optimization (PSO), and k-means. We included these results in this paper, and the comparisons for the image segmentation dataset are presented in Table 3.

Based on the results reported in Table 3, our proposed algorithm has the smallest classification error compared to all other algorithms. If we look at individual efficiency, we can see that the GA, BA, and the proposed BBFWA perform worse for class 3 comparing to the k-means and PSO algorithm. On the other hand, BA and the BBFWA obtained better efficiency for class 1 compared to all other methods. In all other cases, performances of the BA and BBFWA were slightly better in comparison with other methods, the GA and PSO had the similar efficiency while k-means had the worst results which were expected since it found a local best solution for initially randomly chosen centers. The proposed BBFWA achieved better results in most cases.

5 Conclusion

Bare bone fireworks algorithm was adjusted and combined with the k-means clustering algorithm for multispectral satellite image classification. The BBFWA was used for finding the cluster n centers where n is the number of classes. The proposed method performs better in terms of classification accuracy compared to other methods from the literature: original k-means, PSO, GA, BA, MLP BPNN, SVM, and k-NN. In future work, feature selection and different distance metrics can be also included.

References

1. Kelcey, J., Lucieer, A.: Sensor correction of a 6-band multispectral imaging sensor for UAV remote sensing. *Remote Sens.* **4**, 1462–1493 (2012)
2. Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordan, A.: Soil mapping, classification, and pedologic modeling: history and future directions. *Geoderma* **264**, 256–274 (2016)
3. Ngo, L.T., Mai, D.S., Pedrycz, W.: Semi-supervising interval type-2 fuzzy c-means clustering with spatial information for multi-spectral satellite image classification and change detection. *Comput. Geosci.* **83**, 1–16 (2015)
4. Romero, A., Gatta, C., Camps-Valls, G.: Unsupervised deep feature extraction for remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **54**, 1349–1362 (2016)
5. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **55**, 645–657 (2017)
6. Al-Harbi, S.H., Rayward-Smith, V.J.: Adapting k-means for supervised clustering. *Appl. Intell.* **24**, 219–226 (2006)
7. Banerjee, B., Bovolo, F., Bhattacharya, A., Bruzzone, L., Chaudhuri, S., Mohan, B.K.: A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy. *IEEE Geosci. Remote Sens. Lett.* **12**, 741–745 (2015)
8. Shi, Y.: Brain storm optimization algorithm. In: Tan, Y., Shi, Y., Chai, Y., Wang, G. (eds.) *Advances in Swarm Intelligence, LNCS*, vol. 6728, pp. 303–309. Springer, Heidelberg (2011)
9. Dolicanin, E., Fetahovic, I., Tuba, E., Capor-Hrosik, R., Tuba, M.: Unmanned combat aerial vehicle path planning by brain storm optimization algorithm. *Stud. Inf. Control* **27**, 15–24 (2018)
10. Yang, X.S.: Firefly algorithms for multimodal optimization. *Stoch. Algorithms Found. Appl. LNCS* **5792**, 169–178 (2009)
11. Tuba, E., Tuba, M., Beko, M.: Two stage wireless sensor node localization using firefly algorithm. In: *Smart Trends in Systems, Security and Sustainability, LNNS*, vol. 18, pp. 113–120. Springer, Heidelberg (2018)
12. Tuba, E., Mrkela, L., Tuba, M.: Support vector machine parameter tuning using firefly algorithm. In: *26th International Conference Radioelektronika*, pp. 413–418. IEEE (2016)
13. Tuba, E., Alihodzic, A., Tuba, M.: Multilevel image thresholding using elephant herding optimization algorithm. In: *14th International Conference on Engineering of Modern Electric Systems (EMES)*, pp. 240–243 (2017)
14. Tuba, E., Ribic, I., Capor-Hrosik, R., Tuba, M.: Support vector machine optimized by elephant herding algorithm for erythemato-squamous diseases detection. *Proc. Comput. Sci.* **122**, 916–923 (2017)

15. Karaboga, D., Ozturk, C.: A novel clustering approach: artificial bee colony (abc) algorithm. *Appl. Soft Comput.* **11**, 652–657 (2011)
16. Senthilnath, J., Kulkarni, S., Benediktsson, J.A., Yang, X.S.: A novel approach for multispectral satellite image classification based on the bat algorithm. *IEEE Geosci. Remote Sens. Lett.* **13**, 599–603 (2016)
17. Tuba, E., Strunberger, I., Bacanin, N., Zivkovic, D., Tuba, M.: Cooperative clustering algorithm based on brain storm optimization and k-means. In: 28th International Conference Radioelektronika, pp. 1–5. IEEE (2018)
18. Senthilnath, J., Omkar, S., Mani, V.: Clustering using firefly algorithm: performance study. *Swarm Evol. Comput.* **1**, 164–171 (2011)
19. Lichman, M.: UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences (2013). <http://archive.ics.uci.edu/ml>
20. Ghosh, S., Biswas, S., Sarkar, D., Sarkar, P.P.: A tutorial on different classification techniques for remotely sensed imagery datasets. *Int. J. Smart Comput. Rev.* **4**, 34–43 (2014)
21. Li, J., Tan, Y.: The bare bones fireworks algorithm: a minimalist global optimizer. *Appl. Soft Comput.* **62**, 454–462 (2018)
22. Tuba, E., Tuba, M., Beko, M.: Support vector machine parameters optimization by enhanced fireworks algorithm. In: Advances in Swarm Intelligence, LNCS, vol. 9712, pp. 526–534. Springer International Publishing, Cham (2016)
23. Tuba, E., Tuba, M., Dolicanin, E.: Adjusted fireworks algorithm applied to retinal image registration. *Stud. Inf. Control* **26**, 33–42 (2017)
24. Tuba, E., Tuba, M., Simian, D., Jovanovic, R.: JPEG quantization table optimization by guided fireworks algorithm. In: LNCS, International Workshop on Combinatorial Image Analysis, vol. 10256, pp. 294–307. Springer, Cham (2017)

A Survey of Directory and Database Protocols for Data Extraction



Manav Ashok Thakur, Santosh G. Bari, Rutuja Deshmukh and Smita Auty

Abstract In today's world for each and every one, the data is very important factor or data plays a very important role. The concept of computer data is coming from the eighteenth century, and in nineteenth, computer data is very important concept. In nineteenth most of the devices are having the data storage capability. So, this data storage capability leads the computer to invent various protocols. The protocol is the special set of rules that are used to access the data, to exchange the data and to communicate between various nodes of computer. Mainly data is stored in databases or in directory servers in computer. The protocols are also used to compress the data, to notify the data sending device that data sending is over and to notify the receiving data device that data receiving is over. Depending upon the operation, protocols are classified into various types. Protocols are fallen under the Open Systems Interconnection (OSI) models. It consists of various layers and various protocols. Depending upon the use and application, protocols are used.

Keywords Directory · Database · LDAP · ODBC · Schema · SQL · X.500

1 Introduction

Generally looking towards the data storage is that it is nothing but the used to store the data and read the data by integrating various applications to the database or to the directory. The data may be in form of rows and column or stored in hierarchical

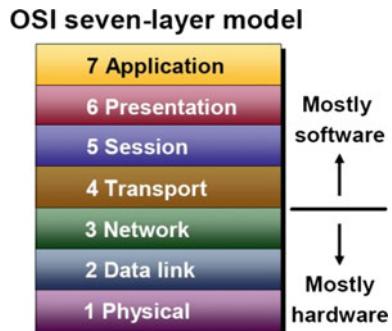
M. A. Thakur (✉) · S. G. Bari · R. Deshmukh · S. Auty
Pune, India

e-mail: manavthakur84@gmail.com

S. G. Bari
e-mail: sgbpune@gmail.com

R. Deshmukh
e-mail: rutuja.deshmukh13@gmail.com

S. Auty
e-mail: smitaauty@gmail.com

Fig. 1 OSI model

structure. Many applications are linked with this database and directory servers to access the data or to query the data [1]. This data transfer or communication is taking place on main networks. In past ages of 1970 to do the communication on networks, the Open Systems Interconnection model (OSI model) is used. This model basically consists of the two major components, and those are abstract model of the networking mainly known as the Basic Reference Model or also called as seven-layer model, and second is a set of specific protocols. But The Open Systems Interconnection model (OSI model) works on the seven layers for the communication [2]. In this ISO model, information is passed from one level to another. This helps protocol working at that point to access information from below working protocol. This shows that each level of protocol is working at given point [2].

- Application: Work as a service provider for application layer
- Presentation: Work as an information converter
- Session: Resolve issues not related to communication
- Transport: Control end-to-end communication
- Network: Help in routing information on network
- Data link: Handle error-related issues
- Physical: Help to transfer entity for media for transmission (Fig. 1).

The need for data communication and security is increased day by day, the new data communication options have come into the market and the people start to work and adopt that.

2 Database

The word database is a collection of information stored in such format that it can be easily accessed, managed and updated, depending upon the requirement. Databases are classified depending upon content type stored in database, and those types are bibliographic, full text, numeric and images. Sometimes databases are also classified according to their organizational approach used by companies. Depending upon this

approach, there are many different kinds of databases, and those are distributed database, relational database, NoSQL database or cloud database [3].

Normally user does not know how the application is connected to database or data is accessed from database. The basic problem behind this is not only applications or softwares which are needed to access databases, but in several conditions, they may access different types of different database. For this purpose, several different technologies, software and protocols are used. While designing any software and protocols to access the database information general approach is consider and depends on that the protocol that access protocols are designed in such manner that, they will try to gather all detail information about available database on one platform without taking specifications in consideration. When any information is requested from client, that time general database protocol will come into picture and it will connect to data source for completing client request. Once information is available, the translation is done in desired language [2, 3]. There are some common protocols that are used to access database.

2.1 ODBC (Open Database Connectivity)

ODBC it is called as application programming interface (API) for accessing information from a database. It is most commonly used database protocol called as Open Database Connectivity (ODBC). This protocol is joined venture of program IBM and Microsoft including other companies came in use in 1980. The core component of ODBC is known as ODBC drivers. These ODBC drivers can be found or bought for virtually for all database platforms. This Open Database Connectivity protocol help to handle client request related to file from different database like Text files, Excel files and Base files.

2.2 Java Database Connectivity (JDBC)

It is used to connect to popular database using for application programming interface which is written in Java language. This API helps to perform all operation on database using Structured Query Language (SQL) statements and also provide the output in same language. Java Database Connectivity protocol is similar to Open Database Connectivity which falls under SQL Access Groups. Using JDBC API, we can connect to any database management system (DBMS) [4]. JDBC includes four different components:

- The JDBC API
- JDBC driver manager
- JDBC test suite
- JDBC-ODBC bridge.

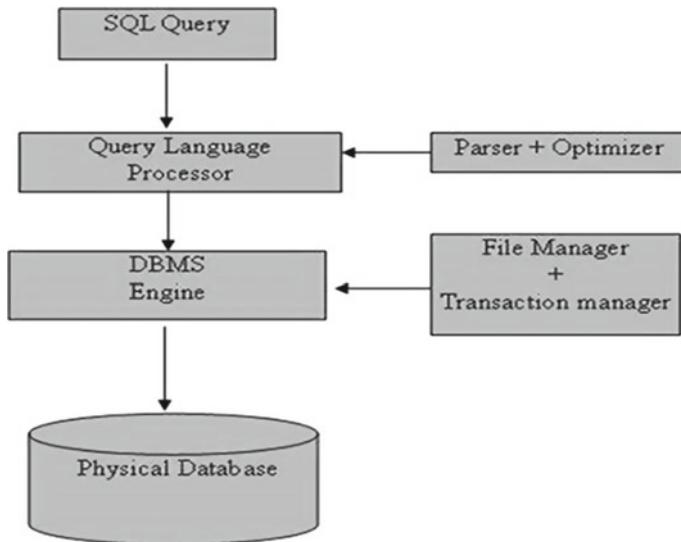


Fig. 2 SQL process

2.3 *SQL (Structured Query Language)*

It is an ANSI/ISO standard. There are different versions of Structured Query Language. It provides various commands to retrieve and manipulate data from database like SELECT, DELETE, UPDATE, INSERT, WHERE. It is mostly used for relational database system such as Sybase, MySQL, MS Access, Oracle, Informix, SQL Server and Postgres to use SQL as their standard database language [5, 6] (Fig. 2).

2.4 *NoSQL*

It is new category of database management systems. It is called as “Not only SQL”. This concept is used by Google, Facebook and Amazon. Due to relational database use, many times system goes down, so to overcome this problem NoSQL is mostly used. That is reason so-called as non-relational databases.

3 Directory

Approach looking towards the directory, It is nothing but the collection or list of data stored in the form of hierarchical structure namely called as Directory Information Tree (DIT). Directory is one type of database repository which is used to store

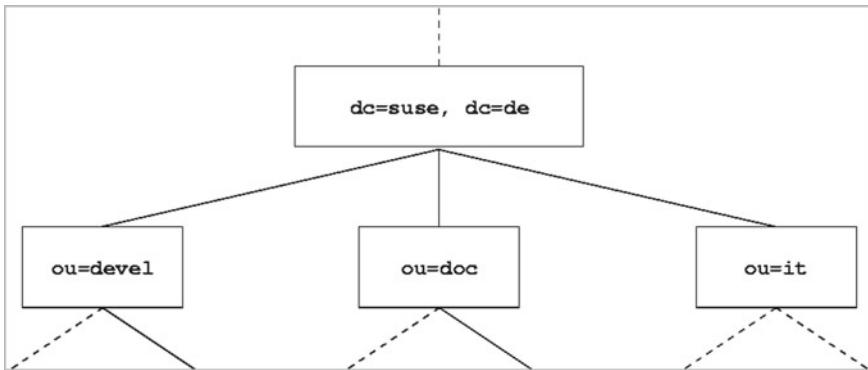


Fig. 3 Directory structure

and manage the information related to user identity profiles and user access privileges, computers, people applications and network devices. It helps people to find information which is in the form of index or list. When the user wants to access any information from the directory server, he has to take the help of protocol that is called as directory protocol [2, 7]. The directory protocol that uses Open Systems Interconnection model (OSI model) is Directory Access Protocol (DAP), Directory System Protocol (DSP), Directory Information Shadowing Protocol (DISP), Directory Operational Bindings Management Protocol (DOP), and the LDAP is the protocol which works on TCP/IP that is why called as Lightweight Directory Access Protocol [8]. This LDAP protocol is mostly used protocol when wants to access any data from directory, that is why it is having an alternative name as Directory Access Protocol. LDAP is the Internet Protocol which is used to access directory servers data (Fig. 3).

3.1 Introduction to X.500

X.500 is originated from the ISO/ITU X.500. X.500 is the directory service protocol and provides standard way to develop an electronic directory of users in an organization and that directory can be accessed using the Internet and also available to anyone in world globally. X.500 comes under the computer networking standards and helps to cover electronic directory service. The X.500 protocol is based on Open Systems Interconnection model (OSI model). It uses some components to access the information from directory and those are an information model, a namespace, a functional model and an authentication framework. The structure of X.500 is different from information storage and retrieval. Information access by X.500 is in the form of hierarchical format. The core concept of the X.500 is there is single Directory Information Tree (DIT) with the hierarchical organization structure of entries which are available in different servers. In the X.500, an entry is to store or identify with

Fig. 4 DN and RDN structure

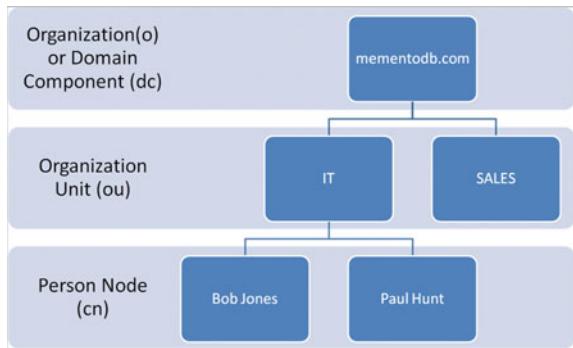
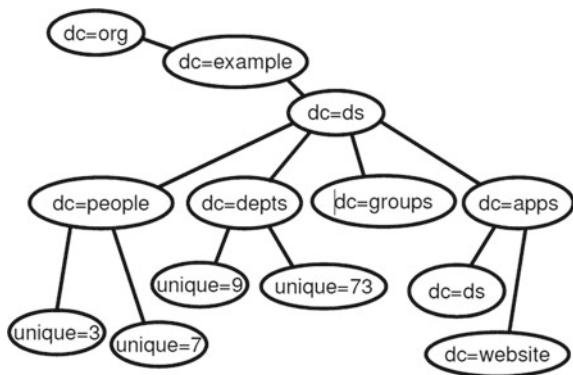


Fig. 5 DIT overview



distinguished name (DN) [7]. The distinguished name (DN) is the combination or we can call as the combination of the relative distinguished name (RDN). So, in other language we can call as distinguished name (DN) that is combination of one or more relative distinguished name (RDN) (Fig. 4).

In each Directory Information Tree (DIT), distinguished name (DN) is unique. It will describe the content of attributes in tree.

The protocols used to access the data from directory are

- Directory Access Protocol (DAP)
- Directory System Protocol (DSP)
- Directory Information Shadowing Protocol (DISP)
- Directory Operational Bindings Management Protocol (DOP) (Fig. 5).

3.1.1 Directory Access Protocol (DAP)

It is the computer network standard used to access X.500 Directory Service. X.500 defines the Directory Access Protocol (DAP) is used by the client to complete the required information from directory. Directory Access Protocol gives any idea how

the query is issued and processed by X.500 Directory User Agent (DUA) that communicates with help of Directory System Agent (DSA). If the users have the suitable permission access then user can view, modify, delete and search the data stored in X.500 Directory using Directory Access Protocol (DAP). Directory Access Protocol (DAP) uses OSI model for its communication that is why called as complex protocol with lot of overhead [2, 7]. This nature makes him unsuitable for in Microsoft Windows environment.

3.1.2 Directory System Protocol (DSP)

The operation of Directory System Protocol (DSP) is similar to Directory Access Protocol (DAP). DSP protocol uses OSI layer for communication, especially application layer. It uses ACSE for establishing and releasing association, and ROSE uses for carrying protocol interaction. The operations allow by the DSP to perform on directory is as similar to DAP but some more operations allow and those are Chained Read, Chained Search and Chained Modify.

3.1.3 Directory Information Shadowing Protocol (DISP)

CA Technologies in 1993 implement the Directory Information Shadowing Protocol (DISP) which allows replicating information stored into directory using OSI model. This replication implementation model supports

- DISP routing
- Shared configuration
- On-demand and periodic updates.

Using DIS Protocol X.500 implement Master-Slave Replication terminology or it can be called as Directory Information Shadowing Protocol (DISP) is used to implement standard replication terminology of X.500.

3.1.4 Directory Operational Bindings Management Protocol (DOP)

The fourth protocol that falls under the category of X.500 is DOP. It is mainly used to make management operational binding in distributed operations. Directory Operational Bindings Management Protocol (DOP) helps to perform operations related to modification, establishment and termination of operational binding.

3.2 Difference Between Database and LDAP Directory

- (1) Directory stores the information in static format.

- (2) Database stores information in dynamic format.
- (3) Directory is mostly used for reading operation.
- (4) Database is used for both reading and writing purpose.
- (5) Directory information is organized in a hierarchical structure.
- (6) Database information is organized in rows and column format.
- (7) LDAP protocol is used to access the information from directory.
- (8) SQL language is used to access the data from database.

4 Lightweight Directory Access Protocol (LDAP)

Few years ago, the company name as Netscape Communications adopted Lightweight Directory Access Protocol (LDAP) with help of different other 40 companies. It is open industry standard to access the directory data in hierarchical format known as Directory Information Tree (DIT). Its name itself Lightweight Directory Access Protocol (LDAP) gives an idea that it is lightweight version of the Directory Access Protocol (DAP) and heavy X.500 protocol [7]. In other language, it can be stated as LDAP is an Internet alternative protocol to X.500 Directory Access Protocol (DAP). The most use of this protocol is in medium-to-large-scale organizations. Lightweight Directory Access Protocol (LDAP) working is completely based on TCP/IP protocol [8]. Due to this, it plays an important role in development of Intranet and Internet application which provide the information about system users, network, services and applications those are using network. Even if the directory is running in different operating system and in different environment, LDAP protocol allows a program called as a browser or an e-mail package to perform operation on directory.

4.1 *LDAP Working Model*

The working operation of LDAP is completely based upon the client–server working model. Applications that want to access the data from Directory or read–write the data from directory do not have direct access to directory server. When any request comes it calls a function or application programming interface (API), which causes LDAP client to send the message to LDAP server. For this communication, LDAP uses the TCP/IP protocol. Then LDAP server accesses the information from directory, and then it sends reply to LDAP client [8, 9] (Fig. 6).

4.1.1 **LDAP Server**

The LDAP server is that which access the information from directory. Server stores the access data locally. LDAP server provides authentication and searching services

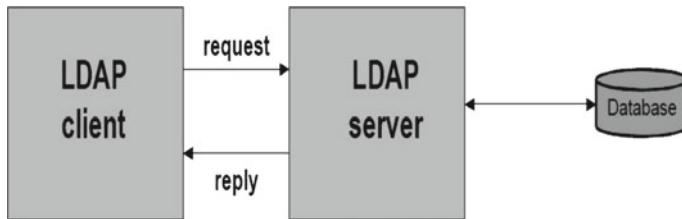


Fig. 6 LDAP client–server model

and operations like add, delete and modify. It provides the access or performs the request of client. When LDAP Client authentication done with LDAP Server that process called as binding with server.

4.1.2 LDAP Client

LDAP client accesses the LDAP server over the network. When LDAP client sends the request or connects with LDAP server that time that is called binding operation. Once binding with server is done, actual operations are performed and reply is sent back to LDAP client. Once LDAP client request is performed, then connection with of LDAP client and LDAP server is broken that is called unbinding with server.

4.2 Replication Using LDAP

Replication is a powerful tool which is used to duplicate data between multiple directories. It helps to improve the performance, scalability and redundancy of data between the directories. The concept behind the replication is to make the directory data available in more than one location by keeping them in synchronization. In LDAP replication terminology, there is a supplier is the server that sends the changes to the other server [9, 10]. And the server that receives the changes from another server called as a consumer. LDAP replication topologies are

- The Master-Replica topology
- The Peer-Peer topology
- The Master-Forwarder-Replica topology.

Below is the replication implementation of replication topology on LDAP server. Server 1 is running on port number 1389, and server 2 is running on port 2389 (Fig. 7).

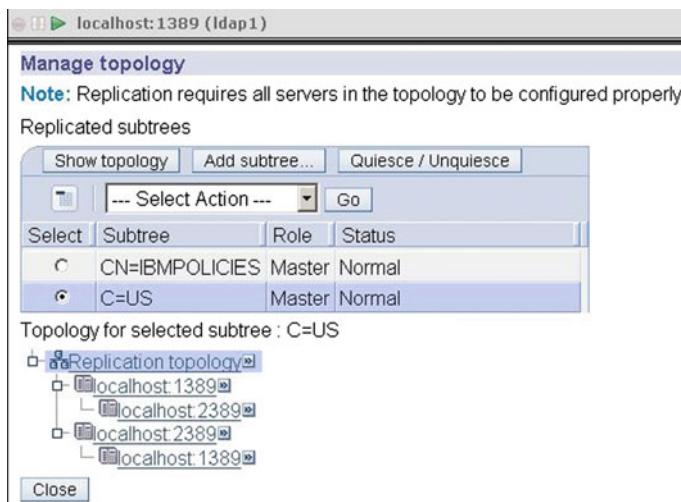


Fig. 7 Replication topology

5 Conclusions

The directory and database protocols are mainly used to access the data from database or directory. These protocols are having different used and operations that used to retrieve the data. All these protocols are supported by different platforms. But each protocol is bounded with either with database or directory. All database protocols are using JDBC connectors, and directory is using DAP. But the purpose behind both is nothing to store the data and retrieve data with fast speed.

References

1. X.500 and DAP Security: A Comparative Overview. IEEE Network, November/December 1999
2. Li, Y., Li, D., Cui, W., Zhang, R.: Research Based on OSI Model
3. Liu, M., Shan, R.: Design and Implementation of the Relation Log Deductive Database System
4. Xia, Q., Feng, L., Lu, H.: Supporting Web-based Database Application Development
5. Sahri, S.: Design of a Scalable Distributed Database System: SD-SQL Server
6. Litwin, W., Rich, T., Schwarz, Th.: Architecture for a scalable distributed DBSs application to SQL Server 2000. In: 2nd International Workshop on Cooperative Internet Computing (CIC 2002), August 2002
7. Hassler, V.: Tech. Univ. Wien, Austria "X.500 and LDAP security: a comparative overview. In: IEEE Network, Nov/Dec 1999
8. Wahl, M., Howes, T., Kille, S.: Lightweight Directory Access Protocol (v3). Google Scholar

9. Salim, M., Sana Akhtar, M., Qadeer, M.A.: Data Retrieval and Security Using Lightweight Directory Access Protocol
10. Patil, R., Zawar, M.: Improving replication results through directory server data replication. In: IEEE Conferences, 22 February 2018

Recent Trends in ICT-Enabled Renewable Energy Systems



Shamkumar Chavan and Mahesh Chavan

Abstract Information and communication technology has wide applications; nowadays, the focus is on use of ICT services in multiple areas for easy and rapid communications. Renewable energy systems are in greater demand, and experimentation is going on in optimization, efficiency improvement, development of fault-tolerant architectures, etc. Recently, a trend is seen to use ICT-enabled services in renewable energy applications. In this manuscript, an overview of ICT-enabled services in renewable energy applications is presented. ICT-based facility makes easy towards monitoring the plant conditions, faults, abnormalities, end applications, etc. Advancements in IoT will be proved to be cheap, easy and user-friendly way for remote power plant monitoring.

Keywords ICT in renewable energy · ICT in PV systems · ICT in fault diagnosis · ICT-based monitoring

1 Introduction

Renewable energy power stations are located at remote sites. Condition, parameter monitoring of power stations is important issue. Nowadays, ICT-based techniques are being implemented everywhere for monitoring various parameters remotely [1]. Several attempts have been done to implement ICT for remote monitoring of renewable energy stations. In this context, this manuscript presents an overview of recent trends in ICT-based monitoring techniques for renewable energy stations. It is seen that web-based, Zigbee-based, GPRS-based and GSM-based monitoring techniques are popular and provide facility of remote monitoring, property analysis, abnormalities detections in power plants .

S. Chavan (✉)

Department of Technology, Shivaji University, Kolhapur, Maharashtra, India
e-mail: sbc_tech@unishivaji.ac.in

M. Chavan

KIT'S College of Engineering, Kolhapur, India

2 Selected Trends in Renewable Energy Monitoring Systems

In this section, an overview of ICT-enabled renewable energy power stations is presented.

Pereira et al. [2] developed monitoring system for distributed renewable energy stations. According to authors for small power plants, implementation of monitoring system may not be cost-effective; therefore, low-cost DAS, transmission system and open-source tools are used. Raspberry Pi is preferred on site, and through cloud servers, the data can be accessed by multiple users. For sensing PV voltage, resistive divider network is used, current is sensed using ACS 712 module, LM 35 is used for temperature sensing, pyranometer LP02 is used for irradiance, and DHT 11 is used for relative humidity. The data is displayed on web page. An IoT-based system is developed [3] for batteries' performance monitoring in smart grids. In this system, battery parameters, viz. voltage, current, temperature, state of the battery charge, are monitored and displayed via GUI on remote device. The data is monitored after every minute of time interval. PV station monitoring system is developed [4] which provides facilities like parameter monitoring, preplanned inspections, data analysis, fault monitoring, etc. In this, SIM 900 GPRS module is used for communication between server and PV parameter data logger. The information can be accessed remotely via computers. Remote monitoring system for maintenance and analysis of power generation system is implemented [5]. In these open-source IoT platforms, LoRa network is used. An IoT-based monitoring system is developed [6] in which sensing circuits are attached to solar PV panels. The system is capable to monitor the parameters, to optimize power under shading conditions, to detect and to correct the faults, to minimize the transients in inverters, etc. Thus, system parameter monitoring and partial controlling are facilitated here. Development of an IoT-based PV monitoring system using Raspberry Pi and MQTT protocol for gateway and phones is discussed [7]. Phung et al. [8] developed IoT-based control system for monitoring PV parameters, meteorological information and energy control in microgrid. Secured PV monitoring system [9] based on SUN which uses AES and ARIA algorithms is described. Periodic performance monitoring of 1.4 MW PV in grid-tied system is described [10].

Cloud-based PV system monitoring and management is discussed [11]. An IoT-based virtual system in LabVIEW for monitoring the condition of PV panels is discussed [12]. The parameters like power, current, voltage, relevant environmental parameters and faults in PV panel are monitored using LabVIEW and are uploaded on cloud. Variation of parameters outside the predetermined levels indicates the faults in PV panel. An IoT-based system for monitoring electrical parameters of load [13] is described. Relay is used as a protecting device. PV panel cleaning and monitoring using IoT are facilitated remotely [14]. This eases removal of dust particles from PV panel and enhances the efficiency.

Architecture is proposed for fault monitoring scheme in PV power plants using Web services [15]. In the proposed system, the master controller communicates with

PV panel circuits and converters through CAN bus. The information is stored in server and can be accessed by remote nodes. In the proposed scheme, the present condition and faults occurred on site can be remotely monitored. This facilitates remote fault diagnosis and decision regarding repairing or component replacement. Further, the intervention of human in fault diagnosis is avoided and time is saved. Reliability of PV system is improved [16] by remote monitoring and comparing the PV system parameters. The monitored information is stored in server and is accessible via web page. Zigbee-based system for monitoring the PV environmental parameters is discussed [17]. This enables the user for stability and efficiency determination, fault detection and notification, sensor failure detection, etc. Data can be seen remotely on web page. WSN-based system for run-time abnormalities' detection in PV panels is discussed [18]. At PV monitor station, SVM is implemented for detecting abnormalities in PV panels. LabVIEW-based system is developed in which PV parameters and weather parameters are monitored and displayed on PC [19]. GUI is designed to graphically monitor all parameters. Meliones et al. developed online monitoring and controlling system for PV power plant [20]. In this system, site information can be accessed by computers and/or cell phones. Critical situations are informed via emails and SMS, facility of generation of reports having information of financial conditions, maintenance and daily operation is also provided via emails and SMS. The data is stored locally and have facility to access it in xml and graphic formats. GSM-based system for condition monitoring of stand-alone system is implemented [21]. Changes in monitored parameters are informed through SMS. Krauter et al. developed monitoring system [22] for remotely located PV power stations in which data is transferred to satellite; the satellite collects the information and sends it to ground station through which data is sent to multiple accessing nodes via Internet. The Brazilian ARGOS-SCD satellite is used for this purpose. Yaqub et al. [23] simulated controlling of desalination plant based on renewable energy in CISCO's packet tracer software. Web access is also provided for the users. Cost-based comparisons of different remote monitoring schemes along with limitations are discussed [24]. In the implemented system, voice channel from GSM network is used to transfer the signal. Xiaoli et al. [25] developed Zigbee-based system for PV, inverter and environmental parameter monitoring. An application of satellite-based system for discrimination of snow coverage on PV panels with other type of abnormalities is discussed [26]. Meghdadi et al. developed [27] snow detection and alarming system for PV panels. In this system Arduino microcontroller is used which monitors irradiance, PV panel current and voltage, when PV panel electrical parameters and irradiance level lowers, the microcontroller send message over internet and owner is informed via Twitter regarding the snow presence.

The overview presents that much of the work is concentrated on monitoring PV or wind panel electrical parameters, fault diagnosis, system failures, environmental parameter observations and end application monitoring.

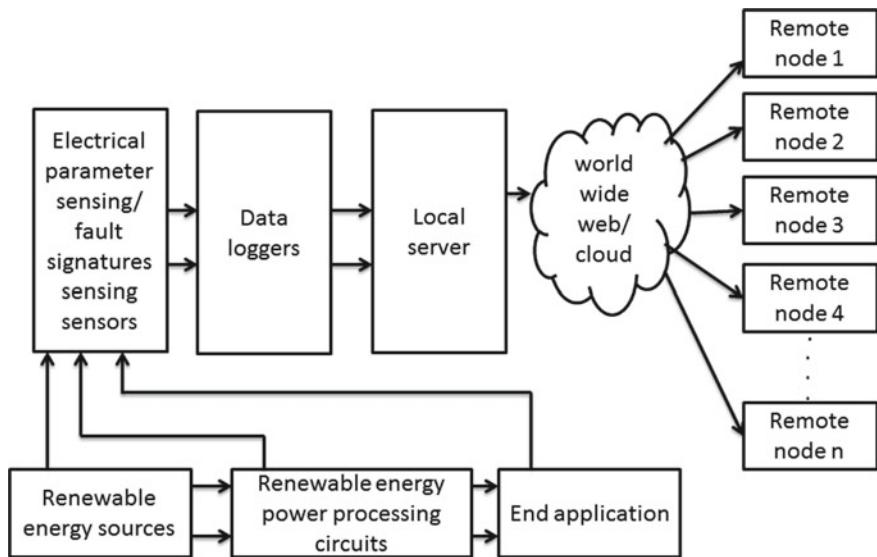


Fig. 1 Typical ICT-based monitoring system for renewable energy application

3 Discussions

Applications of ICT-based techniques in renewable energy power stations are limited for status and fault monitoring purpose. However, this facility enables continuous monitoring from distant place. Generalized ICT-based system for renewable energy application is shown in Fig. 1. PV panels, wind turbines, end application, etc., can be monitored using suitable sensors. Data loggers perform role of digitization, conversions, storing, recording, etc. Data is transferred and stored in local servers in suitable formats. Data loggers transfer data via wires using suitable protocol; wireless data transmission can also be implemented. Remote nodes can access the data using Internet or cloud facility. ICT-based systems make remote condition monitoring, controlling, fault monitoring an easy task and are useful for appropriate decision-making in case of malfunctioning.

Satellite-based monitoring is also implemented in which data is transferred via remote earth stations.

4 Conclusions

ICT-based monitoring schemes for renewable energy power plants facilitate regular condition monitoring remotely as renewable power plants are located at remote and distant sites. Any abnormalities, faults, etc., can be detected and diagnosed,

and appropriate action can be planned remotely to save the time. Daily data can be recorded and made available to study the changing characteristics of systems. IoT-based monitoring is cheap and easy to implement. Evolution in ICT-based techniques and recent inventions in IoT technology will bring user-friendliness in remote monitoring.

Acknowledgements Authors are thankful to Department of Technology, Shivaji University, Kolhapur, Maharashtra, India, for providing necessary facilities for completion of this work.

References

1. Chavhan, S.B., Kadam, P.A., Sawant, S.R.: Embedded web server for monitoring environmental parameters. *Instrum. Exp. Tech.* **52**, 784–786 (2009)
2. Pereira, R., Dupont, I., Carvalho, P., Juca, S.: IoT embedded linux system based on Raspberry Pi applied to real time cloud monitoring of a decentralized photovoltaic plant. *Measurement* **114**, 286–297 (2018)
3. Friansa, K., Haq, I.N., Santi, B.M., Kurniadi, D., Leksono, E., Yuliarto, B.: Development of battery monitoring system in smart microgrid based on Internet of Things (IoT), EPIC 2016. *Proc. Eng.* **170**, 482–487 (2017)
4. Adhya, S., Saha, D., Das, A., Jana, J., Saha, H.: An IoT based smart solar photovoltaic remote monitoring and control unit. In: 2nd International Conference on Control, Instrumentation, Energy and Communication, pp. 432–436 (2016)
5. Choi, C.S., Jeong, J.D., Lee, I.W., Park, W.K.: LoRa based renewable energy monitoring system with open IoT platform. In: International Conference on Electronics, Information and Communication (2018)
6. Spanias, A.: Solar energy management as an Internet of Things (IoT) application. In: International Conference on Information, Intelligence, Systems and Applications (IISA) (2017)
7. Choi, C.S., Jeong, J.D., Han, J., Park, W.K., Lee, I.W.: Implementation of IoT based PV monitoring system with message queuing telemetry transfer protocol and smart utility network. In: International Conference on Information and Communication Technology Convergence, pp. 1077–1079 (2017)
8. Phung, M.D., Villefromoy, M.D.L., Ha, Q.: Management of solar energy in microgrids using IoT based dependable control. In: International Conference on Electrical Machines and Systems (2017)
9. Choi, C.S., Jeong, J.D., Han, J., Park, W.K., Lee, I.W.: Implementation of secured photovoltaic monitoring system on the smart utility network. In: International Conference on Information and Communication Technology Convergence (2016)
10. Aditya, S., Sarwat, A.I.: Roadmap to prepare distribution grid tied photovoltaic site data for performance monitoring. In: International Conference on Big Data, IoT and Data Science, pp. 110–115 (2017)
11. Lee, J., Shin, Y., Lee, I.: Cloud based application platform for smart monitoring & management of photovoltaic generation systems. In: International Conference on Electronics, Information and Communications (2018)
12. Laxmi Prasanna, J., Lavanya, D., Anil Kumar, T.: Condition monitoring of a virtual solar system using IoT. In: International Conference on Communication and Electronic Systems, pp. 286–290 (2017)
13. Pramono, S.H., Sari, S.N., Maulana, E.: Internet based monitoring and protection on PV smart grid system. In: International Conference on Sustainable Information Engineering and Technology, pp. 448–453 (2017)

14. Tranca, D.C., Rosner, D., Palacean, A.V.: Autonomous flexible low power industrial IoT controller for solar panels cleaning systems. In: International Conference on Control Systems and Computer Science, pp. 106–112 (2017)
15. Chavan, S., Chavan, M.: Web based condition and fault monitoring scheme for remote PV power generation station. In: Information and Communication Technology for Sustainable Development, Lecture Notes in Networks and Systems, vol. 10, pp. 133–141. Springer, Singapore (2018)
16. Ventura, C., Tina, G.M.: Utility scale photovoltaic plant indices and models for on-line monitoring and fault detection purposes. *Electr. Power Syst. Res.* **136**, 43–56 (2016)
17. Shariff, F., Rahim, N.A., Ping, H.W.: Zigbee based data acquisition system for online monitoring of grid connected photovoltaic system. *Expert Syst. Appl.* **42**, 1730–1742 (2015)
18. Hu, T., Zheng, M., Tan, J., Zhu, L., Miao, W.: Intelligent photovoltaic monitoring based on solar irradiance big data and wireless sensor networks. *Ad Hoc Netw.* **35**, 127–136 (2015)
19. Rohit, A., Tomar, A., Kumar, A., Rangnekar, S.: Virtual lab based real time data acquisition, measurement and monitoring platform for solar photovoltaic module. *Resour. Efficient Technol.* **3**, 446–451 (2017)
20. Meliones, A., Apostolacos, S., Nouvaki, A.: A web based three tier control and monitoring application for integrated facility management of photovoltaic systems. *Appl. Comput. Inf.* **10**, 14–37 (2014)
21. Gagliarducci, M., Lampasi, D.A., Podesta, L.: GSM based monitoring and control of photovoltaic power generation. *Measurement* **40**, 314–321 (2007)
22. Krauter, S., Depping, T.: Remote PV system monitored via satellite. *Sol. Energy Mater. Sol. Cells* **82**, 139–150 (2004)
23. Yaqub, U., Al-Nasser, A., Sheltami, T.: Implementation of a hybrid wind solar desalination plant from an Internet of Things (IoT) perspective on a network simulation tool. *Appl. Comput. Inf.* (in press)
24. Tejwani, R., Kumar, G., Solanki, C.: Remote monitoring of solar photovoltaic systems in rural application using GSM voice channel. In: ISES Solar World Congress 2013, Energy Procedia, vol. 57, pp. 1526–1535 (2014)
25. Xiaoli, X., Huan, W.: The wireless network construction of the photovoltaic power system-Zigbee. *Proc. Eng.* **15**, 2511–2515 (2011)
26. Wirth, G., Homscheidt, M., Zehner, M., Becker, G.: Satellite based snow identification and its impact on monitoring photovoltaic systems. *Sol. Energy* **84**, 215–226 (2010)
27. Meghdadi, S., Iqbal, T.: A low cost method of snow detection on solar panels and sending alerts. *J. Clean Energy Technol.* **3**, 393–397 (2015)

Assessing the Smart Readiness of Local Councils in Mauritius



**B. Gobin-Rahimbux, N. Gooda Sahib-Kaudeer, N. Chooramun,
S. Cheerkoot-Jalim, M. Heenaye-Mamode Khan, Z. Cadersaib, S. Kishnah
and S. Elaheeboocus**

Abstract Mauritius, like most countries in the world, is cherishing the dream of becoming smart, and the Government of Mauritius is already providing incentives in order to embark on a smart island initiative. Existing local councils will play a major role in this transformation process. In this work, the smart readiness of local councils is assessed based on their respective departments and the services they deliver to the local population. Focus groups were carried out with representatives from local council departments to identify the current status of their services, the level of technology involvement and the extent to which services contribute to smartness. The challenges faced by local councils to improve their smart readiness were also discussed. The smart readiness assessment revealed that some departments are more ready than others. In some departments, there is a significant gap in the current delivery of services. The prospect of becoming smart requires the services to become

B. Gobin-Rahimbux (✉) · N. Gooda Sahib-Kaudeer · N. Chooramun · S. Cheerkoot-Jalim ·
M. Heenaye-Mamode Khan · Z. Cadersaib · S. Kishnah · S. Elaheeboocus

Faculty of Information Communication and Digital Technologies, University of Mauritius,
Réduit, Moka, Mauritius

e-mail: b.gobin@uom.ac.mu

N. Gooda Sahib-Kaudeer
e-mail: n.goodasahib@uom.ac.mu

N. Chooramun
e-mail: n.chooramun@uom.ac.mu

S. Cheerkoot-Jalim
e-mail: s.cheerkoot@uom.ac.mu

M. Heenaye-Mamode Khan
e-mail: m.mamodekhan@uom.ac.mu

Z. Cadersaib
e-mail: z.cadersaib@uom.ac.mu

S. Kishnah
e-mail: s.kishnah@uom.ac.mu

S. Elaheeboocus
e-mail: shabneez.elahheeboocus@umail.uom.ac.mu

computerised, intuitive and efficient. The assessment in this work therefore elicits the level of effort required to transform local councils, which in turn will form an integral part of the smart island initiative of the country.

Keywords Sustainable development · Smart readiness assessment · Smart island

1 Introduction

Since recent years, the Government of Mauritius is placing a lot of emphasis on the development of smart cities, as a means to boost its economy and to improve the standard of living of its citizens. The government planned to establish eight smart cities, with three of those already in development [1–3]. The development of these smart cities will be primarily based on the Greenfield model, thereby applying the top-down planning methodology [4]. This model will incorporate intelligence, technology and smartness in the service delivery right from the inception and early construction of smart cities. However, the worldwide trend is the transformation of existing cities into smart ones using the Brownfield model. Conversely to new developments, a bottom-up regeneration approach [4] is required to introduce intelligence, technology and smartness in existing cities. The transformation of these cities presents several challenges in terms of design and implementation strategies, due to their existing, persistent and complex problems. In order to be successful in its smart island initiative, the Government of Mauritius is envisaging the inclusion of smartness in the existing urban and rural areas. Local governance is ensured by district councils (rural areas) and municipal councils (urban areas), which will be jointly referred to as local councils in this work.

Before embarking on transformation projects for councils in Mauritius, it is fundamental to investigate the extent to which the local councils are ready to become smart. The work presents the assessment of the current status of each service delivered, the technology involvement and the contribution of these services to smartness. It presents the scores obtained following a focus group carried out with representatives from various departments of local authorities. The average score for the smart readiness of each department of the local councils is presented. These results are expected to help understand the current situation and help in setting up of an ICT architecture for smart local councils in Mauritius.

2 Background Study

2.1 *Smart Mauritius*

The concept of smart cities is still new and definitions vary [5–8]. According to the White Paper prepared by the smart cities Project Team of IEC [9], there are three main pillars of development for any city, namely economic, social and environmental sustainability. Smart cities focus on the use of digital technologies to provide better services to their citizens by making better use of resources. The concept of smart cities goes beyond the relationship between the service provider (mainly the government) and the service users (the citizens). Smart cities provide for platforms to enable citizens to be more participative in the community for instance, by providing feedback on the quality of services being offered. In the same breadth, a smart city does not refer to a final phase in its development process but it is rather a series of processes or steps which make cities better to live in and be more responsive to forthcoming challenges thereby resulting in happy citizens.

The smart approach should also be adopted by Small Island Developing States. According to [10], a smart island is an insular territory that can create a sustainable economy for its people while ensuring a high quality of life. Infrastructure, natural resources and the environment should be managed by smart, innovative and integrated solutions to promote a governance approach that is socially inclusive and to tackle challenges such as climate change. Islands have their own challenges and particularities. Though islands may consider the same activities/pillars/dimensions as smart cities in Europe or Asia, the way these activities/pillars/dimensions are implemented is not necessarily similar. For instance, in Mauritius, some services that would normally be handled by local government in a Smart city are being handled on national level. This is discussed in details in [11]. The role of local authorities and the services they provide will differ from those services provided by smart cities.

Smart governance in Mauritius will involve both central government and local councils. At present, most of the computerisation has been done at the level of the central government. In [12], the importance of structured governance has been highlighted for the purpose of designing and implementing policies with regard to city and district developments. In another study by [13], governance was found to be conducive to the coordination and integration of departments. Therefore, to be successful in a smart island initiative, local councils will play a determinant role and they will have to embark on smart governance. But before going ahead with such an endeavour and having smart local governance, it is of utmost importance to carry out a smart readiness assessment within local councils. Such an activity had not been done with respect to Mauritius.

2.2 Smart Readiness Assessment

Smart city readiness assessment has been undertaken in several studies. For instance, [14] explores different smart city initiatives and existing smart city transformation frameworks in order to assess the readiness of the Arab region to achieve smart city projects, as a challenge to transform its cities into smart sustainable cities. Many cities in the region have already embarked on smart city initiatives; however, a few cities are still suffering from a knowledge gap related to the transformation, mainly because of a set of constraints at the city or country level, namely socio-economic inequality, political instability, inappropriate ICT infrastructure and lack of well-managed transportation network [14]. However, the assessment tool used in this study was not presented.

Furthermore, PwC [15] investigated the extent of the use of technology in city administration and the levels of service delivery to citizens in the major cities of India. In this work, a holistic transformation strategy was devised to address economic, social and environmental issues and to take advantage of forthcoming opportunities. The first phase of this strategy included an assessment of the current status of various sectors contributing to city operations, expectations of all major stakeholders as well as the readiness for transformation. This phase included a systematic analysis of key sectors, namely municipal corporation, health and water management, and transport amongst others. Three performance measures, namely current status, technology intervention and contribution to smart city, were used to compute the readiness scores for each sector [15].

In [16], the smart city development readiness of cities in Vietnam was assessed using an expert perception survey approach. In this study, participants from both governmental and non-governmental organisations, who are conversant with smart city development concepts, were surveyed with regard to city development policies in their respective cities. The survey examined different dimensions including smart governance, smart economy, smart human capital development, smart infrastructure and mobility and smart environment [16]. The readiness scores pertaining to all the dimensions investigated were mentioned; however, the calculation leading to those scores was not presented.

Following the review of these three studies, the assessment framework presented by PwC was adopted due to the fact that the performance measures on which the smart readiness assessment was based are clearly defined and explained. In the context of the local councils in Mauritius, it was possible to use similar performance measures to assess the smart readiness of the services, department and ultimately the local councils. In Sect. 3.2, the performance measures are redefined according to the local councils in Mauritius and the approach used to assess of smart readiness is described.

3 Assessing the Smartness of Local Councils in Mauritius

Local councils in Mauritius are governed by the Ministry of Local Government in accordance with the Local Government Act 2011 [17]. According to this law, local councils have to provide services such as scavenging/refuse collection, cleaning and maintenance of open spaces, roadwork, drain works, managing of markets and fairs amongst others. In Mauritius, there are one city council, four municipal councils and seven district councils. Local councils in Mauritius have six departments, namely Administration, Finance, Public Infrastructure, Land Use and Planning, Welfare and Public Health. Different councils have adopted different modes of operations and are not uniform. Till date, many services are paper based. Examples include the application for supplier registration and that of the building and land permit.

A focus group meeting was organised to find out the extent to which technology is being used in the day-to-day operations of the local councils and their readiness to adopt smart initiatives. In this focus group meeting, the city council, the four municipal councils and six district councils have participated.

3.1 *Methodology*

Different focus groups were organised for each department. The purpose of regrouping the same department of all councils was to identify the similarities and differences in service delivery, current technology use and the challenges faced. There were eleven participants each from Administration and Finance, seven from Land, Use and Planning, nine from Public Health, six from Public Infrastructure and three from Welfare. In order to capture both qualitative and quantitative information, a systematic approach was used. Prior to the focus group, a questionnaire was designed. The questionnaires were sent to the participants before the focus group so that they have an insight into the topics discussed. Each focus group had a facilitator, a presenter and at least two observers for recording the discussion. The presenter explained the purpose of the meeting to the participants and used a presentation to guide each participant to fill in the questionnaire. A thematic analysis of focus group notes was carried out by the observers to ensure accuracy, validity and reliability of the responses recorded. The observers of each session compiled the recordings. During this process, any query or missing information was validated with the respective council department. The data compiled per department was then verified with the presenter and facilitator. The scores for each service were calculated, and the qualitative findings of each department were compiled. Some of the findings are presented in Sect. 4.

Each service provided by the local councils was assessed for their readiness to become smart, and thus, a Smart Readiness Score (SRS) was calculated. The assessment was based on three components, namely current status (RS), technology involvement (RT) and contribution to smartness (RC) as proposed by [15]. For the

Table 1 Definition of the components used to assess smart readiness

Component	Definition
Current status (RS)	Defines how efficiently the service is currently being delivered
Technology involvement (RT)	Defines the extent to which technology is being used to deliver the service
Contribution to smartness (RC)	Defines how important the service is in determining the smartness of the local council

purpose of assessing the local councils, these components were redefined to encompass the context in which they are being used (Table 1).

Calculating Smart Readiness Score. Each component was assessed on a scale of 1–3, with 3 being the highest rating. For component RC, it was observed that representatives from different local councils perceived the contribution of the service to smartness differently, despite the service and its importance being the same for all local councils. Most representatives viewed the services in their respective departments as highly contributing to smartness as shown in Fig. 1, Panel C. Therefore, to prevent any kind of bias for component RC, the RC ratings for each service by different councils were averaged before being used in Eq. (1). For the remaining two components, the average score was not computed as different councils deliver the same service differently. For example, for the same service, local council A might still be using a fully manual system, whereas local council B is using a computerised system. Therefore, RS and RT for this service for both A and B will be different. Additionally, weights have been assigned to each component based on their significance and their contribution to the smart readiness assessment of the councils. The weights assigned in this paper differ from [15] in order to be relevant to the local context. RS and RT were given an equal weightage of 0.4 as there are greater possible variations to be accounted for. RC was assigned a weight of 0.2, which represents at least half of the weightage of RS and RT, since municipal councils already provide targeted services for a specific geographical location. Thus, the total score for each service (TS) is calculated as follows:

$$TS = (RS \times 0.4) + (RT \times 0.4) + (RC_{avg} \times 0.2) \quad (1)$$

where

RS Score for indicating how efficiently the service is currently being delivered.

RT Score for indicating the extent to which technology is being used to deliver the service.

RCavg Average rating for RC across all local councils for a service

Subsequently, the departmental score, represented by $Score_{dept}$, is obtained by adding the total score for each service in that particular department. For the purpose of

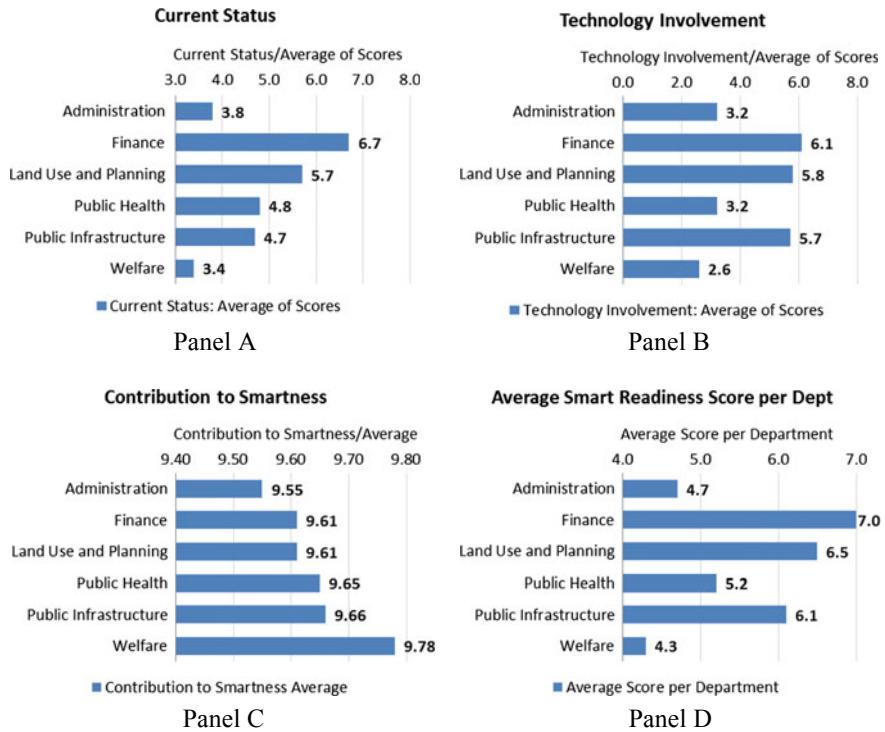


Fig. 1 Average scores per department

standardisation and comparison, a score index for each department is also calculated and is set in a range of 0–10 based on Eq. (2):

$$\text{ScoreIndex}_{\text{dept}} = \frac{\text{Score}_{\text{dept}}}{\text{BestScore}_{\text{dept}}} \times 10 \quad (2)$$

where

$\text{Score}_{\text{dept}}$ Score obtained by one department in one local council.

$\text{BestScore}_{\text{dept}}$ Best possible score that can be obtained by one department

Thus, a department scoring an index nearing 10 will indicate a high readiness to become smart while a score nearing zero will highlight the need for significant efforts to become smart.

Table 2 Average score for each department across all local councils

Department	Average of ScoreIndex _{dept}	Range (Score _{indexdept})
Administration	4.7	(4.0–5.5)
Finance	7.0	(6.8–7.5)
Land use and planning	6.5	(5.8–7.0)
Public health	5.2	(4.6–6.7)
Public infrastructure	6.1	(5.2–7.1)
Welfare	4.3	(3.8–5.1)

4 Findings and Discussion

This section discusses the main findings of this work. Table 2 shows the average smart readiness score for each department.

Panels A, B and C of Fig. 1 illustrate the average scores for current status, the technology involvement, contribution to smartness per department. Panel D shows the average smart readiness score after computation as per Eq. (1).

4.1 Administration

The Administration department in local councils focuses on internal matters related to the council. For example, it deals with all external communication, human resource management of council staff and organisation of council meetings. This department has a low smart readiness with a score of 4.7, and this is largely due to the scores for current status and technology involvement which are at 3.8 and 3.2, respectively. Qualitative findings from the focus groups reveal that many services such as leave management and recruitment within this department across local councils are still manually delivered. A few local councils have attempted to implement biometric attendance record systems, but faced resistance from staff. In other cases, the current efficiency of services has been queried as there are no general guidelines to dictate how local councils use technology for the same services. For example, for document management, no standardised software has been used. Oracle Financials is used for any procurement and finance-related activity as well as Microsoft Outlook in some cases. Internet access is essential for a high level of technology involvement especially for a centralised approach across different councils. All staff did not have Internet access, and some not even a personal computer.

4.2 Finance

The Finance department has achieved a high score of 7.0 and thus is more ready to become smart compared to other departments. The current status of services in the department, as well as the technology involvement, contribute to this high score significantly as they are both high at 6.7 and 6.1, respectively. During the focus groups, participants revealed that this department is computerised across all councils and they are already using several technology tools such as E-Business Suite from Oracle Financials. There are some services where payment is done manually. For example, citizens who use the cremation facilities after business hours or on weekends have to effect payment directly to the worker who then has to wait until the next business day in order for the payment to be processed. Such circumstances could be catered for with mobile applications to allow citizens to make payments at any time of the day. Services such as payment of trade fee are already available online. However, participants highlighted that many citizens are not using the online payment services as it is not user-friendly, and there is no awareness campaign to encourage them to use these online services. The different councils are planning to implement the online payment for other services as well and to conduct sensitisation campaigns to encourage citizens in their adoption.

4.3 Land Use and Planning

The Land Use and Planning department of local councils allocates building permits, monitors the development of projects for the councils and handles complaints related to unauthorised constructions. Across all local councils, this department is the second most ready to become smart with a score of 6.5. Participants highlighted that this department already uses technology for some tasks. For example, applicants can use their credit cards to pay their processing fee online, and building applications are scanned and archived. Staff use email for communication. In addition, Geographical Information Systems (GIS) are used to obtain satellite views of localities under their jurisdiction in order to monitor unlawful expansion and construction projects. However, there are still some tasks that are not computerised which therefore hinder the progress towards smart readiness. For instance, there are no electronic means for the application, processing and delivery of building permits. Local councils face many difficulties in service delivery, as it is partially online. For example, local councils receive complaints both through the online Citizen Support Unit and through walk-in complaints desks and letters. While the council inspectors work with soft copies of the scanned building plans, they have to inform applicants of decisions through registered posts. This often results in manual archives as well as duplication of work especially when different ministries are involved.

4.4 Public Health

The Public Health department is responsible for services such as waste collection, market fairs, rodent control, general cleaning of rivers and public infrastructure. This department has a smart readiness score of 5.2, which highlights the need for significant improvement to become smart. The two components affecting this overall score are the current status and technology involvement which have a low score of 4.8 and 3.2, respectively. This indicates that services need to be computerised and optimised. Participants in the focus groups described some initiatives to use technology. However, this was not the case for all councils as these efforts are the results of the staff's own initiatives rather than official guidelines. While some councils incorporate a degree of smartness in their activities, other councils are still far behind in this endeavour although some facilities are available. For example, some councils are using Global Positioning Systems (GPS) and GIS to monitor waste collection lorries while others use WhatsApp to send pictures to health inspectors while monitoring unoccupied lands and for solid waste management. It was observed that the lack of IT usage in some cases was not necessarily due to the unavailability of the related tools. Other factors include the lack of training and the staff's reluctance to use these tools. In order to improve smart readiness, business processes such as bidding, contract allocation and management, allocation of market stalls, monitoring of cemetery and cremation grounds need to be computerised.

4.5 Public Infrastructure

The Public Infrastructure department handles the implementation of projects related to roads, drains, lighting amongst others. It currently has a smart readiness score of 6.1 with its current status at 4.7 and technology involvement at 5.7. The current status score indicates that services within this department should become more efficient in order to make service delivery smart. Participants in the focus groups discussed that project management is an important part of their work. Some councils are still recording project details manually, some use spreadsheets while a few others use project management software. Staff in this department is aware of the benefits of smart approaches such as use of solar panels, sensor-based monitoring of drains and simulation tools amongst others. Lack of technical know-how and funds were the main challenges identified.

4.6 Welfare

The Welfare department has the lowest smart readiness score of 4.3, across all local councils. This is reflected in its score for current status and technology involvement

which are at 3.4 and 2.6, respectively. The services under this department include the provision and management of gymnasium facilities, sports complex, swimming pools and libraries. These services vary amongst the local councils, but the booking process is done manually. Local council websites are used for advertisement, but very little information is available and the websites are not regularly updated. This department is one where local citizens are major stakeholders as they regularly interact with the councils with regard to the services provided. Therefore, in order to improve smart readiness, the services provided by the councils have to be user centric. A few participants highlighted the use of word processors, spreadsheets and email. However, in order to become smart, there is a need to develop applications to improve the quality of service in councils. Such applications include mobile applications to citizens for booking of services and getting instant notifications amongst others.

5 Conclusion and Future Works

This work focused on assessing the smart readiness of local councils in Mauritius. Every service offered by each department under their respective local councils was assigned a score using three performance measures, namely current status, technology involvement and contribution to smartness. Scores were calculated from ratings provided by representatives of the local councils through focus groups. The assessment highlights that some departments in local councils have a low smart readiness score. The main reasons are lack of efficiency in the delivery of the service, low involvement of technology, reluctance to use existing computerised systems and lack of training and funds. It is believed that if the delivery of those services is re-engineered and the appropriate technology is used, local councils can provide smart services, and thus contributing to the smart island initiative. To further support this endeavour, the future works proposed are as follows:

- An in-depth needs analysis of citizens: The needs analysis is essential because the quality of services currently being delivered to citizens needs to be assessed. The main focus should be on the level of user-centeredness of these services.
- Technology acceptance survey: To investigate the readiness of the citizens towards adopting technology for using the local council services.
- Identification of new smart services to be provided to the population: This will be performed based on the needs analysis and the current state of the art in smart local governance. It is understood that the services identified in the literature will need to be adapted with respect to the local context and national priorities.
- Design of a smart ICT architecture for local councils: To design a blueprint for future implementation and transition towards smartness.

Acknowledgements We are thankful to the Mauritius Research Council for funding this project as well as all the participants of the local councils of Mauritius for the valuable input provided.

References

1. Budget Speech 2015–2016: <http://www.investmauritius.com/media/343854/BudgetSpeech-2015-2016.pdf>. Last accessed 2018/06/11
2. Budget Speech 2017–2018: http://budget.mof.govmu.org/budget2017-18/2017_18budgetspeech.pdf. Last accessed: 2018/06/11
3. Ackbarally N.: The making of smart Mauritius. In: African Business Magazine. <https://africanbusinessmagazine.com/region/southern-africa/the-making-of-smart-mauritius/>. Last accessed 2018/06/11
4. Komninos, N., Tsarchopoulos, P., Kakderi, C.: New services design for smart cities: a planning roadmap for user-driven innovation. In: Proceedings of the 2014 ACM International Workshop on Wireless and Mobile Technologies for Smart Cities, pp. 29–38. ACM (2014)
5. Kunzmann, K.R.: Smart cities: A new paradigm of urban development. In: Academia, https://www.academia.edu/9530213/SMART_CITIES_A_NEW_PARADIGM_OF_URBAN_DEVELOPMENT. July 2014, last accessed 2016/01/10
6. Ludlow J., Khan, Z.: Participatory democracy and the governance of smart cities. In: 26th Annual AESOP Congress, Ankara, Turkey (2012)
7. Khan, Z., Kiani, S.L.: A cloud-based architecture for citizen services in smart cities. In: EEE/ACM Fifth International Conference on Utility and Cloud Computing (UCC'12), pp. 315–320 (2012)
8. Caragliu, A., et al: Comparative performance assessment of smart cities around the North Sea basin comparative performance assessment of smart cities around the North Sea basin. Netw. Ind. Q. **13** (2011)
9. IEC: Orchestrating Infrastructure for Sustainable Smart Cities. White Paper (2014)
10. Smart Island Initiative: <http://www.smartslandsinitiative.eu/en/index.php>. last accessed 2018/06/14
11. Sahib-Kaudeer, N.G., Devak, S., Gobin, B.A.: What is ‘Smart’ for smart island developing states. In: IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (2016)
12. Kumar, H., Singh, M.K., Gupta, M.P., Madaan, J.: Moving towards smart cities: solutions that lead to the smart city transformation framework. Technol. Forecast. Soc. Change (2018)
13. Lee, J.H., Hancock, M.G., Hu, M.C.: Towards an effective framework for building smart cities: lessons from Seoul and San Francisco. Technol. Forecast. Soc. Change, 80–99 (1989)
14. Ibrahim, M., Adams, C., El-Zaart, A.: Paving the way to smart sustainable cities: transformation models and challenges. JISTEM-J. Inf. Syst. Technol. Manag. **12**(3), 559–576 (2015)
15. PriceWaterhouseCoopers: How Smart are our Cities? http://www.pwc.in/assets/pdfs/publications/2015/how_smart_are_our_cities.pdf. Last accessed 2016/01/10
16. Vu, K., Hartley, K.: Promoting smart cities in developing countries: policy insights from Vietnam. In: Telecommunications Policy, pp 1–15 (2017)
17. Local Government Act 2018: Republic of Mauritius Portal. <http://localgovernment.govmu.org/English/Legislations/Pages/Local-Government-Act.aspx>. Last accessed 2018/06/16

The Economics of ‘Make in India’ Over ‘Buy (Import)’ Decision in Selected Technologies for the Indian Navy



Alok Bhagwat and Pradnya Vishwas Chitro

Abstract Indian Navy has embarked upon an ambitious fleet expansion, and modernization programme Indian Prime Minister Narendra Modi during the International Fleet Review at Visakhapatnam, India, in February 2016 underlined the need for a ‘modern and multidimensional Navy’. Indian Navy has already established itself as a ‘Builders’ Navy with a majority of its warships being built indigenously. However, a submarine plays a vital role in sea denial and sea control roles of a nation. Therefore, in pursuance of the role laid down in the Maritime Security Strategy Document, published by the Indian Navy in the year 2015, Indian Navy is also laying equal emphasis on indigenous submarine construction. Indian Navy has planned an indigenous submarine construction programme, christened P75 (I). Indian Navy is driving towards achieving a significant indigenous content in P75 (I), in line with the ‘Make in India’ initiative. Accordingly, it is mandatory for the foreign shipyards to forge a strategic partnership with a select Indian shipyard where the submarines will be built. With an outlay of about INR. 60,000 Cr (USD 8 Billion), this programme is seen as a game changer in indigenous submarine construction. The P75 (I) project throws open significant opportunities in indigenous development of submarine equipment. Today, a majority of submarine equipment are imported. The aim of this paper is to suggest a road map for indigenous manufacture of equipment and systems for P75 (I). The research methodology will involve literature review as secondary methodology, interviews with stakeholders and pilot studies as primary research methodology. It will be an exploratory research. There is a significant cost advantage for the country and Indian industry in indigenous manufacture of this equipment. Further, the indigenously developed equipment is expected to have export potential and applications in civil industry.

A. Bhagwat · P. V. Chitro (✉)
Symbiosis International (Deemed) University, Pune, India
e-mail: pradnyac@sims.edu

A. Bhagwat
e-mail: bhagwat.alok@gmail.com

P. V. Chitro
Symbiosis Institute of Management Studies (SIMS), Pune, India

Keywords Make in India initiative • Indigenization of prime technologies • Life cycle cost

1 Introduction

Seaborne terror and piracy are two key challenges to maritime security. These have resulted in Indian Ocean region remaining Indian Government's priority given India's 1200 island territories, and its huge exclusive economic zone of 2.4 million square kilometres. Indian Prime Minister has further underlined the Indian Ocean region's role 'as a strategic bridge with the nations in our immediate and extended maritime neighbourhood'. Underlining the need for a 'modern and multidimensional Navy', Prime Minister Modi stressed that India would continue to actively pursue and promote its geopolitical, strategic and economic interests on the seas, in particular the Indian Ocean [20].

India must pay greater attention to the modernization and expansion of its naval forces and their ability to meet a wide range of threats. A large part of this maritime modernization policy has another domestic political aspect to it: the concept of the 'Indianization of the Navy'.

The theme of indigenization is central to warship building plans of the Indian Navy which had embarked upon indigenous shipbuilding journey long ago. The indigenization theme of Indian Navy got a strong boost by the 'Make in India' initiative. Indian Navy came out with Indian Naval Indigenization Plan (INIP) 2015–2030. The INIP clearly brings out the areas in which indigenous development is desired [13].

2 Literature Review

IHQ MoD (N) [17] brought out the role of Indian Navy revolving around four main areas as follows:

- **Military.** The Navy's military role is characterized by the threat or use of force at and from the sea. For this, the Navy is required to be prepared for a variety of warfare, e.g. surveillance, anti-submarine warfare (ASW), anti-surface warfare, anti-air warfare (AAW), electronic warfare (EW), mine warfare and harbour defence. In addition, the Navy is required to protect Sea Lanes of Communication (SLOC) and advise to the Indian shipping companies on the choice of safe routes across the seas and on entry/exits from harbours in times of hostilities. Therefore, there is a need for Indian Navy to maintain minimum force levels.
- **Diplomacy.** Naval diplomacy entails the use of naval forces in support of foreign policy objectives to build 'bridges of friendship', strengthen international cooperation, on the one hand, signal capability and intend to deter potential adversaries on

the other. One of the important tasks in the diplomatic role is overseas deployment of ships.

- **Constabulary.** The increasing incidence of maritime crime has brought into sharp focus the constabulary role that Navies have to perform. The significance of this role may be gauged by the fact that for a third of the world's navies, this is a major facet of their functions. Counterterrorism, counter-infiltration and anti-piracy operations are common evolutions for the Indian Navy.
- **Benign.** Examples of the benign tasks include humanitarian aid, Search and Rescue (SAR), disaster relief, ordnance disposal, diving assistance, salvage operations and hydrography survey.

The above roles demand a formidable fleet and need for building more and more ships indigenously. Further, in addition to indigenous ship design and construction, the equipment package is required to be manufactured indigenously.

Kennedy and Pant [20] analysed the Maritime Security Strategy Document and found out the desire on the part of Indian Navy to modernize and increase its warfighting capabilities across all aspects of naval capability. They found out that the document makes a powerful and thoughtful case for why India must pay greater attention to modernization and expansion of its naval forces. A large part of the maritime modernization policy has another domestic political aspect to it: the concept of the 'Indigenization of the Navy'. The authors also found out that the theme of indigenization runs throughout the statement.

Chandramohan [11] found out that India which followed continental strategic orientation since independence has now laid greater attention to its maritime capabilities. They noted that India is strengthening its Navy and extending the range of naval operations. Consequently, the countries that are wary of China's increased maritime assertiveness most recently displayed in the South China Sea dispute are likely to move towards strategic partnership with India.

Sakhuja [28] found out the broader maritime aspects of Indo-US cooperation. He argues that India has also announced national initiatives and action plans to promote Blue Economy. The priority sectors for India's Blue Economy ecosystem include shipping, ports, SEZ and shipbuilding.

Mohanty [24] found out that the Indian Navy is slowly yet determinedly taking to the blue waters. He argues that the service's current vision is to become a competent, confident, operationally viable and regionally visible maritime power, and to slowly expand its stabilizing role in and beyond the Indian Ocean. He highlights that due to both political/strategic and economic considerations, since the early 1960s, the Indian Navy is progressively reducing its dependence on foreign shipyards and is aiming at self-sufficiency. He notes that Indian Navy has formulated a very ambitious 15-year plan for an expansion to 185 ships by 2017.

Kanwal [19] found out Indian Navy's ambitious plan of expanding by 150 ships in next ten to fifteen years. Accordingly, Indian Navy has about 50 ships on order and about 100 more in acquisition pipeline.

Tellis [31] found out that India and USA have agreed to form a working group to explore the joint development of India's next-generation aircraft carrier.

Pant and Khurana [27] found out that India has featured on the American radar as an important strategic partner. The authors suggest that USA should provide the Indian Navy technological assistance in such key projects as the nuclear propulsion and the design and construction of aircraft carrier. GlobalSecurity.org [15, 16] talks about modernization plans of Indian Navy. It clearly brings out the opportunity for Indian defence manufacturers. It states that the Indian Navy's efforts will supplement those of other government agencies, as also the private sector. The Science and Technology Roadmap (2015–2035) and Indian Naval Indigenization Plan (2015–2027) will steer the efforts to synergize overall research and development efforts in the maritime sector.

Singh [29] takes a stock of the Scorpene and Arihant submarine construction programmes of the Indian Navy. He makes a case of the need for indigenization in submarine building programmes of the Indian Navy.

Singh [30] brings out the enthusiasm created globally by the announcement of Rs. 64,000 Cr indigenous submarine programme.

Indian Defence News [18] has underlined the keenness of Russia to support indigenous conventional submarine building programme of India.

Buckingham and Hodge [10] found out about the merits of the conventional propulsion system of a submarine. They also found out the need of an air-independent propulsion (AIP) system for submarines. They analysed the performance parameters of bit types of propulsion and found out that with AIP, a submarine is able to remain under water for longer durations.

Behra [5] explains the offset policy of a number of nations and compares the same with India's defence offset policy. He suggests way ahead for a robust offset policy for India.

MoD [23] promulgated a booklet by, compiled by the Department of Defence Production in the year 2016, post Make in India. This brings out provisions made under Make in India initiative in terms of defence offsets, defence exports, industrial licensing and foreign investment.

MoD [22] laid down rules and procedures for defence procurement.

Dhawan and Sahane [12] brought out the relevance of the integrated electric propulsion (IEP) to Indian industry. They suggested a brief road map for Indian industry to gear up for manufacturing IEP equipment within the country.

Thomas and Mohanty [32] found out India's ambitious journey to become a world-class manufacturer of defence products under Prime Minister Narendra Modi's flagship 'Make in India' initiative. They have highlighted that over the next seven years, India plans to build a vibrant manufacturing industry that can not only cater to its armed forces but also begin exports.

BMI Research [9] reports activities of defence PSUs giving insight into five years forecast of defence procurement.

Behra [4] has traced the overall performance of the Defence Industrial Base (DIB) consisting of 52 research laboratories, 9 PSUs and 39 odd ordnance boards and rated it as sub-optimal. He argues that the sub-optimal performance has resulted in the country spending billions of dollars annually on imports. Underlining the need to improve self-reliance in defence technologies, he looks into various aspects of India's

attempt towards self-reliance and brings out recommendations towards achievement of the same.

Behal [3] found out Indian naval warship building programmes till the year 2027 and highlighted the financial outlay planned for the same. She has highlighted that the order book of government defence shipyards stands at Rs. 86,660 crores and that of private shipyards at Rs. 6300 crores. She brings out a 10% YoY increase in naval capital budget, expected to take it to Rs. 70,000 Cr by the year 2026–27. Highlighting the emphasis laid by Defence Ministry on indigenization, she points to the Indian Naval Indigenization Plan (INIP) 2015–2030, promulgated by Ministry of Defence in the year 2015. She explains that the Indigenization Plan particularly focuses on opening up opportunities for private sector and inviting the Indian industry to produce locally what the Navy has been importing.

DoI [13] lists down the areas of indigenization where Indian Navy has desired import substitution by encouraging Indian industry to set up in-country manufacturing set-up.

Bhagwat and Chitrapur [6] have traced the history of naval shipbuilding and growing indigenous content in it. In the light of 'Make in India' initiative, there is a significant scope for Indian defence manufacturers in contributing towards this sector.

Arshad [1] defined life cycle cost of a ship as the total cost of ownership of a ship and its equipment. He elaborated that the LCC of a ship includes its conception, acquisition, operation, maintenance, upgrades and decommissioning. Accordingly, LCC is normally calculated by summing up the cost estimates from inception to disposal or in other words cost estimates from cradle to grave. He went on to argue that while LCC analysis is carried out for commercial ships to estimate the net profit and return on investment, in the case of naval shipbuilding, it is mainly used to choose the most cost-effective options from all the options available.

Barringer [2] argued that the LCC economic model provides better assessment of long-term cost-effectiveness of projects that can be obtained with only first cost decisions. He further explained that business must summarize LCC results in net present value (NPV) format considering depreciation, taxes and time value of money. He went on to justify that LCC helps change provincial perspectives for business issues with emphasis on enhancing economic competitiveness by working for the lowest long-term cost of ownership which is not an easy answer to obtain. He further listed down life cycle costing process by breaking down into 11 steps and their interdependence. He then applied the steps to analyse the three alternatives of a pump operation to elaborate the application of steps.

Langdon [21] carried out a literature review of life cycle costing and life cycle assessment. It mainly focused on civil construction projects in European Union and aimed at to assess the relevance and impact of life cycle costing towards sustainable construction. It covers various aspects of life cycle costing in detail and is expected to contribute towards formulation of life cycle model in the proposed research study.

Bhagwat and Chitrapur [7] have advocated adaptation of collaborative approach towards indigenous development of Electromagnetic Aircraft Launch System (EMALS) for Indian Navy.

Bhagwat and Chitrap [8] found out the ambitious Indigenous submarine construction programme of the Indian Navy highlighting the technology areas for indigenous development (pp 3).

Arshad [1] defined life cycle cost of a ship as the total cost of ownership of a ship and its equipment. He elaborated that the LCC of a ship includes its conception, acquisition, operation, maintenance, upgrades and decommissioning. Accordingly, LCC is normally calculated by summing up the cost estimates from inception to disposal or in other words cost estimates from cradle to grave. He went on to argue that while LCC analysis is carried out for commercial ships to estimate the net profit and return on investment, in the case of naval shipbuilding, it is mainly used to choose the most cost-effective options from all the options available.

Langdon [21] carried out a literature review of life cycle costing and life cycle assessment. It mainly focused on civil construction projects in European Union and aimed at to assess the relevance and impact of life cycle costing towards sustainable construction. It covers various aspects of life cycle costing in detail and is expected to contribute towards formulation of life cycle model in the proposed research study.

Pagoropoulos et al. [25] analysed the costs and benefits of product-service system (PSS) as servitized offerings and its influence on formulation of service strategies in the shipping industry. They used life cycle costing (LCC) as the tool for assessing the costs and benefits for two study areas.

Pande [26] argues that in view of the limited financial resources and, consequently low budgetary allocations for defence, it would be necessary to factor in life cycle cost in the selection of defence equipment. He goes on to state that the decision to procure a military aircraft must not be based on the initial procurement cost alone.

El-Akruti et al. [14] aim to develop maintenance policies for critical assets to improve the production performance based on life cycle cost (LCC) analysis. They define the LCC structure for decision process involving determination of optimum life, repair limit and selection of materials, and trade-off between repair and replacement.

3 The Research Gap

The brief literature review clearly brings out the need for a Blue Ocean Navy as foreseen in Navy's Maritime Doctrine. It brings out the modernization and fleet expansion plans of the Indian Navy and the current shipbuilding orders under execution in the defence and private shipyards. It is obvious that the Indian Navy has an ambitious shipbuilding programme for induction of a large number of ships and submarines including specialized vessels like aircraft carrier and nuclear submarines. The Indian Navy has indigenized ship design and construction to a great extent and is by and large self-sufficient in these areas. A number of ship equipment are also being supplied from within the country. However, there is a large gap as regards indigenous manufacture of a number of critical technology areas in naval shipbuilding. The 'Make in India' initiative is sure to encourage and incentivize Indian industry for establishing

manufacturing capacity within the country. A large capital outlay has been planned by the Navy in building new vessels in the next fifteen years. Indian industry can benefit by associating with indigenous development of the critical technologies that are still being imported for the naval warship building.

There is, thus, a strong case in laying down a robust process to establish successful, profitable and sustainable manufacturing set-up in country for these technologies. The life cycle costing finds a mention in DPP-2016, mandating the OEM to provide the estimated life cycle cost of the product and the basis thereof. The procurement agencies of the Navy, therefore, should consider the life cycle cost for award of a contract. But this practice is not established as yet as there is a gap in estimation of the life cycle cost. Therefore, till date a majority of contracts are awarded on the basis of acquisition cost. The supplier is liable for equipment performance only up to warranty period. As per the DPP-2016, the following factors are to be considered for arriving at the life cycle cost

- Operational hours per year
- MTBF
- Requirement of maintenance spares
- Mandatory replacements during preventive maintenance schedules, etc.

However, as mentioned above, the concept of life cycle costing is never applied while deciding the award of contracts where an Indian supplier is competing with a foreign OEM. Thus, there is a need to study and formulate a life cycle costing model for selected technologies in Indian Navy's warship building programme. The proposed study is aiming to bridge this gap by developing a model for estimation of life cycle cost.

4 The Research Problem

Indian Navy has identified technologies for indigenous development in its ships and submarine construction programme. A number of Indian manufacturers are competing with foreign OEMs for getting orders for this equipment. The DPP-2016 mandates that life cycle costs should be considered for deciding the L1 bidder. However, in the absence of a clear model for establishing life cycle cost, the L1 bidder is decided on the basis of acquisition cost alone. This puts the Indian supplier at a great disadvantage. Therefore, there is a need to establish life cycle costing model for technologies being imported by Indian Navy. The following technologies have been identified for the research study:

- Ship Magazine Fire-Fighting System
- Submarine propulsion motor and control system.

These technologies are explained in brief to bring clarity on the scope and significance of these technologies for Indian Navy.

Ship Magazine Fire-Fighting System (MFFS). A naval warship carries a large volume of explosives onboard. These include gun shells, pyro-techniques, missiles, torpedoes, rockets, decoys and depth charge. The ship employs a very elaborate fire detection and fire-fighting system spread throughout the ship to detect and quench the fire in the compartment holding the explosives. Needless to state, the speed of detection and response is of utmost importance as any slippage would result in the ship blowing up itself. Such accidents have happened in naval history, and the information is available in open domain. At present, the system is imported at a huge price. There is a need to develop this system within the country. The technologies involved are mature within the country, viz. sensitive temperature scanning and pressure detection, water quenching at high pressure with suitable nozzles providing large coverage and inert gas system for missile compartments. There is a need for a management approach in bringing various expert groups together and form a consortium. There is no need to get into technical details as the technologies are available. A strategic management approach with elements of financial and operation management is required to cobble together a model that can ensure profitability to industry players and indigenous solution to Indian Navy. Needless to state, Indian Navy will extend all support in this indigenous development. The researcher has selected this as one of the areas of the research study.

Submarine Propulsion Motor and Drives. The primary mode of moving a submarine is using a motor driving a propeller. On the contrary, surface ships move with the help of diesel engines that drive a gearbox and a shaft. With the advancements in the motor technology, the surface ships also use propulsion motors as these are observed to be cost effective. Today, Indian Navy is importing the propulsion motors and associated drives. Since a majority of submarines were ‘bought out’ or built under foreign design like the Scorpene project, Indian Navy had to accept motors and drives of foreign make. As such, there is no Indian industry player for catering to Indian Navy in these technology areas. It is a common knowledge that India has adequate technological expertise in motors and drives within country. Therefore, establishing a set-up within country to manufacture motors and drives for Indian Navy is not a technological challenge. It is a management challenge. There is a significant market potential, and Indian Navy’s shipbuilding and submarine construction programmes in the near future make a business sense for establishing an in-country set-up for propulsion motors and associated drives. The researcher has selected this as the second research area, viz. indigenous manufacture of propulsion motors and associated drives for Indian naval ships and submarines. There is a scope for exporting these equipments subsequently.

5 Research Objectives

The research study will therefore seek to explore and investigate whether the identified technology equipment should be bought from foreign countries or made in India.

6 Research Methodology

6.1 Primary Research

- Map the market potential for the selected technologies and the execution cycle to arrive at financial cost in-flow model.
- Analyse life cycle cost elements involved in above technologies in the context of Indian Navy’s warship building and submarine building programmes.
- Interview with Indian Navy, defence shipyards and Indian industry, CII, FICCI, DRDO and Defence Ministry to validate the above models.
- Carry out a cash outflow analysis to arrive at viability and profitability of the project for the Indian manufacturer. Identify global players for outright import of these technologies and estimate life cycle cost.
- Compare the cost to arrive at the competitive option.
- Interact with Indian Navy and MoD for incorporation of life cycle costing as a tool for deciding L1 bidder in forthcoming projects. Interview with Indian industry players in defence sector in shortlisting areas for indigenous development.
- Interview with Indian Navy for validation of the areas identified for indigenous development survey of global technology leaders and interaction with them for selection of suitable partner.
- Identify Indian industry players for forming a consortium for joint development and manufacturing within the country.
- Finalization of a commercial model and workshare arrangement with the consortium player to ensure win-win situation for all.

6.2 Secondary Research

- Study of Defence Procurement Procedure 2016 document and Indian Naval Indigenization Plan document for relevance of ‘Make in India’ concept with stipulations for life cycle costing in focus
- Study of life cycle costing models applicable for foreign advanced navies (USA, Canada, France, etc.) and establish a model suitable to Indian Navy for the selected technologies
- Literature review for identifying commercial potential of the selected technology area.

7 Data Collection and Analysis

The data collection was based on the shipbuilding and submarine construction programme of the Indian Navy. This will be collated with inputs from Indian Navy and leading defence and private shipyards.

8 Contribution of the Research Study

The proposed study will lay down the following:

A life cycle costing model for deciding on the most competitive bidder for complex technologies in Indian Navy
It can be extended to the other two services.

9 Limitations

- Limited literature
- Frequent and unpredictable slippages in the shipbuilding and submarine construction programmes deter prospective Indian players in venturing into defence production areas as the capital investment gets locked up for a long time affecting cash inflows and thus recovery. However, there is a positive change in this aspect as MoD is striving to expedite the projects.

These are the primary limitations of the study. However, the process is becoming more transparent gradually and barring some very sensitive areas, the ministry and naval officials are forthcoming in sharing the relevant details. The researcher has identified three technology areas that are not considered sensitive and the concerned Naval Directorates are keen to get these technologies developed within the country.

As regards the slippages in the project schedules, these hit the industry adversely as the inventory piles up and the investment is not recovered in the envisaged time. However, gradually the Indian Navy is improving on this front and a conservative approach towards revenue generation forecast can mitigate this risk to a great extent.

10 Findings

10.1 Secondary Findings

10.1.1 Problems with Imports

Indian Navy’s thrust on indigenization is well founded. An imported equipment comes with its share of problems. It is observed that the foreign suppliers deliver the equipment and hand it over after completion of acceptance trials. The suppliers’ liability ends here. A technology transfer seldom takes place. Some spares are supplied along with equipment, and the ship’s crew is dependent upon these spares for servicing the equipment and ensuring operational availability of the equipment. Once the spares are exhausted, the equipment cannot be serviced. Any further defects in the system require support of the foreign OEM who exploits this opportunity to the fullest and charges heavily to the Indian Navy. Further, there are instances of support denial by the foreign equipment supplier in scenario of conflicts. It is therefore observed that the operation and maintenance of equipment throughout its service life cost are much more for an imported equipment. But this aspect is never taken into consideration while taking the decision about the most competitive bidder and awarding the contract during tendering process.

To cite an example, Indian Navy contracted construction of two oil tankers to an Italian shipyard, Fincantieri. The ships, INS Deepak and INS Shakti, were delivered to the Indian Navy in the year 2011. The ship is installed with a state-of-the-art control system supplied by M/s Sistema, Italy. Within two years of ship commissioning, the system became non-operational and the OEM demanded a huge amount for deputing technicians for servicing the equipment. Indian Navy was not able to justify such maintenance costs in a new ship. Consequently, the system performance remained sub-optimal for a long time. There are many such examples that highlight the problems with imports.

To summarize the problems with imports, the researcher would like to emphasize that an imported equipment may be priced lower at the time of purchase than an indigenous equipment. But Indian Navy eventually pays much more in the operation and maintenance of the equipment throughout its operational life. On the contrary, the indigenous equipment costs much lesser during the exploitation cycle.

10.1.2 Boost to Indigenization

The theme of indigenization is not new to Indian Navy and Indian industry. Indian industry has been supporting Indian Navy with Indigenous systems for a long time. The Electrical Group of L&T (Larsen & Toubro) has been associated with Indian Navy for more than a decade. In this short journey, this business group has indigenized a number of critical technologies for the Indian Navy [6]. Similarly, Tata Group, Mahindra Group, Bharat Forge, etc., have also contributed towards indigenization

in Indian Navy. However, a methodical approach is missing. The DPP-2016 has generated a significant enthusiasm in the Indian industry involved in manufacturing in defence sector. Such an environment has been sought by major industry players and institutions like FICCI, CII and L&T for a very long time. The DPP-2016 is yet to percolate to the ground level where the Indian industry starts contributing in a big way towards defence manufacturing. However, the stage is set for manufacturing houses to commence restructuring of the internal processes and getting collaborations/joint ventures with global technology leaders in place so as to be ready for fulfilling the trust and confidence posed by the Make in India initiative in them.

In order to holistically compare the competing OEMs for critical technologies, there is a need to establish the concept of Life cycle costing while deciding the so-called L1 bidder, a terminology used for the bidder with the lowest price. The researcher proposes to carry out a study for formulation of life cycle costing model as the basis of arriving at the L1 bidder in certain technologies that are still being imported and are shortlisted by Indian Navy for indigenization. In the opinion of the researcher, the life cycle cost is the correct basis for evaluation of two bidders and deciding upon the lowest bidder.

10.1.3 Life Cycle Costing Model

The life cycle cost of equipment is not clearly defined in Indian naval procurement documents. However, as per the general understanding, the life cycle cost of equipment comprises of the following:

a. **Initial Acquisition Cost**

This involves development, design, manufacturing and user trials. The equipment is then installed onboard a ship and the service life commences. The manufacturer is responsible for defect rectifications in the warranty period, usually one to two years. Thereafter, the liability of the supplier ends. At this stage, the Indian Navy's maintenance agencies assume the responsibility of servicing the equipment. For new technology equipment, the Indian Navy does not possess the requisite expertise to ensure operational availability and is heavily dependent upon the foreign OEM.

b. **Operation and Maintenance during Service Life**

This phase extends up to 25–30 years. During this phase, the Indian Navy expects support from the OEM on cost basis. This involves supply of spares, services and upgrades in the event of an obsolescence. This cost is never taken into consideration during initial acquisition and can alter the commercial equations between the bidders.

Life cycle Cost		
Development Cost	Design and Production Cost	Utilization Cost

Comparison—Acquisition Versus Operation Cost

In case of a ship, the life cycle cost is divided into two parts—the initial cost or the acquisition cost and the future cost or the sustaining cost. The initial cost is incurred prior to purchase of the ship, and the future cost is incurred after the commissioning of the ship. The various elements of the life cycle cost of a ship are depicted below [1].



The same model is, by and large, applicable to equipment also. Some of the critical technology equipments that are still being imported cause the Indian Navy to incur a significant expenditure in future maintenance cost of the equipment. These costs are not considered at the time of award of the contract that is decided on the acquisition cost alone.

At an estimate, the magnitude of these two elements of the life cycle cost is given below [1].



The acquisition cost is about 20–40% of the LCC, and the sustaining cost would be about 60–80% of the LCC.

The proposed study aims to formulate a life cycle costing model for estimation of life cycle cost of selected technologies that are considered critical for Indian Navy's shipbuilding programme. The model can then form as the basis of deciding on the contract award, as is stipulated in DPP-2016.

Comparison—Make in India over Import

The acquisition cost is observed to be low in import as the technology is already mature with the foreign OEM and it does not involve development cost, whereas an Indian supplier will have to invest in development and incur a significant capital expenditure.

However, the high development coat is more than compensated if the life cycle costs are compared as the Indian Navy will not incur heavy costs during equipment operation phase. It is observed that this element of the cost is exorbitant for an import.

Primary Research Findings

The researchers carried out a study creating a dummy project and formulating life cycle costing model for applying in make or buy decision. It is expected that this model can be recommended to MoD and Indian Navy in deciding the L1 bidder for certain technologies. This model will give a boost to 'Make in India' initiative as more Indian suppliers will win the contracts involving technologies identified in Indian Naval Indigenization Plan.

Pilot study was conducted by carrying out a market survey identifying various global technology leaders in the areas of IBS. The following were contacted, and initial discussions conducted.

1. M/s OSI, Canada
2. M/s Lockheed Martin, USA
3. M/s MECSPB, St. Petersburg, Russia

The pilot study has laid down the road map for the development of indigenous integrated bridge system (IBS) for the Indian Navy. Till date, there is no fully indigenous supplier of the IBS for Indian Navy. Therefore, there is a lack of self-reliance in this area. A number of technological advancements in the ship control systems are held up as this vital piece of shipborne technology is not available indigenously. The formation of a consortium between an Indian manufacturer and M/s MECSPB has resulted in the development of an indigenous IBS. This development is scheduled to be completed in one year. This will enable indigenous life cycle support of IBS as well as further developmental work in ship control system areas. At the same time, the commercial interest of M/s MECSPB has also been ensured, thus paving a long-term association between an Indian firm and a global technology leader. Indian manufacturers can adapt this model to indigenize critical technologies into Indian Navy. Therefore, the study is of importance in laying down road map for other technologies selected for research study.

11 Conclusions

Indian Navy has achieved indigenous design and warship building capability catapulting the country in a league of select few nations capable of building aircraft carrier and nuclear submarine. However, a lot is to be achieved in in-country manufacturing of submarine equipment. Make in India initiative has evoked enthusiastic response in taking up indigenous manufacturing of high-technology equipment within the country. DPP-2017 has laid emphasis on providing level playing field to the private sector in defence manufacturing. The logical step now is in achieving global collaboration for creating capability within the country in critical defence technology areas so as to help achieving life cycle support from within the country. It makes financial sense as well. Collaborative approach is the only way forward in this direction as has been demonstrated by the EAIC Marine Business of L&T in indigenization of ship control system and magnetic stealth equipment for the Indian Navy. Indigenous development of submarine technologies in consonance with the built schedule of the P75 (I) project following globalization model is the need of the hour. It will help the country in achieving total indigenization in the critical technologies of submarine building. The technological offshoots of the research study will have impact in a number of technology areas. Last but not least, the developed technologies have a significant export potential.

12 Recommendations

Indian industry should undertake urgent capability building in the areas of technologies for supporting P75 (I) programme indigenously. It may need establishing a consortium of Indian industry, research agencies and academia, duly supported by a Global Technology Leader. The technologies should be developed in line with the built schedule of P75 (I) submarine programme. It has a significant export potential and will give the country a strategic edge in naval warfare and establishing sea control and sea denial capability across not only Indian Ocean but also Pacific Ocean.

Bibliography

1. Arshad, T.K.: Ship’s life cycle cost. Published in <https://tkays.com/2013/01/31/ships-life-cycle-costing/> (13 Jan 2013)
2. Barrington, H.P.: A life cycle cost summary. In: International Conference of Maintenance Societies (ICOMS—2003). <http://www.barringer1.com/pdf/LifeCycleCostSummary.pdf> (20–23 May 2003)
3. Behal, R.: Naval shipbuilding financials. Published in <http://www.defproac.com/?p=2700> (2015)
4. Behra, L.K.: Indian Defence Industry—Issue of Self Reliance. IDSA Monologue Series No. 31, p. 6 (July 2013)

5. Behra, L. K.: Defence Offsets—International Best Practices and Lessons for India. IDSA Monologue Series No. 45, pp. 86–89 (June 2015)
6. Bhagwat, A., Chitrap, P., Dr.: Make In India—Opportunity for manufacturing industry in Naval Shipbuilding presented in SIMSARC 2016 (pp. 4, 5). IJRCM (Int. J. Res. Comput. Appl. Manage., 43–46. ISSN 2231-1009. (15 Dec 2016)
7. Bhagwat, A., Chitrap, P., Dr.: Transforming India 2030—strategies for sustainable development goals—partnership for indigenous development—collaborative approach for indigenous development of EMALS for Indian Navy. Presented in SIU (Symbiosis International University) seminar (16 Feb 2017)
8. Bhagwat, A., Chitrap, P., Dr.: Global association for indigenous manufacture of critical technology systems in submarine construction program of Indian Navy. Presented at ABRM-AASTMT International Conference on the Issues in Business & Entrepreneurship, Supply Chain Management & Information Systems (ICBESI), Alexandria, Egypt—Globalization and International Trade (23 May 2017)
9. BMI Research: India—Defence & security report—includes five years forecast to 2020. Published in BMI's Industry Report and Forecast Series. <http://store.bmiresearch.com/india-defence-security-report.html> (April 2016)
10. Buckingham, J., Hodge, C.: Submarine power and propulsion—application of technology to deliver customer benefits. Presented at UDT Europe in Glasgow, pp. 1, 4 (2008)
11. Chandramohan, B.: India looks outward. NZ Int. Rev., 17. <http://search.informit.com.au/payPerView;action=addOrderItem;dn=419412106065700;res=IELHSS;type=pdf> (May/June 2013)
12. Dhavan, C., Sahane, C.C.: Integrated full electric propulsion for Indian Navy—challenges, roadmap and expectations. In: FICCI Seminar—Make in India Paradigm—Roadmap for a Future Ready Naval Force, p. 274 (2016)
13. DoI (Directorate of Indigenization) IHQ MoD (N): Indian Naval Indigenization Plan (INIP) 2015–2030, pp. 6–8 (2015)
14. El-Akruti, K., et al.: Developing and optimum maintenance policy by life cycle cost analysis—a case study. Int. J. Prod. Res. **54**(19), 59465–5962. Retrieved from <https://www.tandfonline.com/doi/full/10.1080/00207543.2016.1193244> (11 May 2016)
15. Global Security.Org: P75(I) programme. <http://www.globalsecurity.org/military/world/india/project-75-i.htm> (29 Mar 2016)
16. Global Security.Org: India-Navy modernisation. <http://www.globalsecurity.org/military/world/india/in-navy-development.htm> (28 Nov 2016)
17. IHQ MoD (N) (Integrated Headquarters of the Ministry of Defence (Navy)): Maritime Doctrine Indian Navy naval strategic publication, 1.1 Updated Online revision in year 2015 (p. 92) (2015)
18. Indian Defence News: Russia to strengthen India's submarine fleet. <http://www.indiandefensenews.in/2016/06/russia-to-strengthen-indias-submarine.html> (28 June 2016)
19. Kanwal, G.: India's military modernization: plans and strategic underpinnings; produced by The National Bureau of Asian Research for the Senate India Caucus, p. 6 (22 Sep 2015)
20. Kennedy, G., Pant, H.V.: Indian Navy's maritime security strategy—an assessment published in National Maritime Foundation, pp. 1, 3, 4, 5 (8 Aug 2016)
21. Langdon, D.: Literature Review of Life cycle Costing (LCC) and Life Cycle Assessment (LCA). Published in http://www.tmb.org.tr/arastirma_yayinlar/LCC_Literature_Review_Report.pdf (June 2006)
22. MoD (Ministry of Defence, Department of Defence Production): Indigenous Plans of Defence Shipbuilding, Submarine Building and Production, p. 11 (2016)
23. MoD (Ministry of Defence): Defence Procurement Procedure—2016 (2016)
24. Mohanty, D.R.: Coming of Age—The Indian Navy in the 21st Century, published in Mil-Tech (Military Technology), pp. 93, 96. (<http://connection.ebscohost.com/c/articles/14182972/coming-age-indian-navy-21st-century>) (July 2014)
25. Pagoropoulos, A., et al.: The influence of costs and benefits' analysis on service strategy formulation: learnings from the shipping industry. Cogent Eng. **4**, 1328792 (17 May 2017). Retrieved from <https://www.cogentoa.com/article/10.1080/23311916.2017.1328792>

26. Pande, B.K., Air Marshal (Retd.): Importance of Life-Cycle Cost. Published in SP’s Aviation Issue 10. Retrieved from www.sps-aviation.com (2015)
27. Pant, H.V., Khurana, Y.: The American Pivot and Indian Navy—its hedging all the way. Naval War College Rev. **68**(1), 47, 64 (2015, Winter)
28. Sakhuja, V., Dr.: Blue Economy—Expanding India—US Maritime Cooperation. IDSA (Institute of Defence Studies and Analysis, New Delhi), p. 2 (14 Sep 2016)
29. Singh, A.: India’s Submarine Modernization Plans. IDSA Monologue (5 May 2016)
30. Singh, R.: Rs 64,000-cr project to build submarines in India creates buzz. <https://www.hindustantimes.com/india/rs-64-000-cr-project-to-build-submarines-in-india-creates-buzz/story-hNwPyJUshugN7x38Fv35QJ.html> (29 Mar 2016)
31. Telis, A.J.: Making waves—aiding India’s next generation aircraft carrier. Paper published in Carnegie Endowment for international Peace (pp 218) (22 April 2015)
32. Thomas, B., Mohanty, D.R.: India’s Quest for Indigenization—Indian Defence Procurement Procedure and Planning. MT (Military Technology), p. 34 (2016/2)

Naïve Bayes Classification on Student Placement Data: A Comparative Study of Data Mining Tools



Umang Mavani, Vivian Brian Lobo, Aditi Pednekar,
Naomi Christianne Pereira, Rupesh Mishra and Nazneen Ansari

Abstract Data mining (DM) is used to analyze and classify data and identify hidden patterns stored in a data warehouse in an attempt to predict future trends, which are quintessential to knowledge discovery and provide tremendous support not only to the world of business but also to that of academia. There are various open-source and freely available software tools such as Weka, R, and Orange as well as programming languages like Python used for DM. This study focuses on comparing the performance of these tools by performing Naïve Bayes classification on student placement data. Percentage of marks scored by students in S.S.C. and H.S.C. examinations and their engineering aggregate were inputs to the tools. Moreover, the tools were trained and tested to decide whether a student would be placed or not. Comparative analyses of the tools were done to determine which tool was able to provide the highest prediction accuracy on student placement data.

Keywords Data mining · Naïve Bayes classification · Placement

U. Mavani (✉) · A. Pednekar · N. C. Pereira · R. Mishra

Department of Computer Engineering, St. Francis Institute of Technology, Mumbai, India
e-mail: umangmavani15@gmail.com

A. Pednekar
e-mail: adi410pednekar@gmail.com

N. C. Pereira
e-mail: naomicpereira19@gmail.com

R. Mishra
e-mail: rupesh.mishra@gmail.com

V. B. Lobo · N. Ansari
Department of Information Technology, St. Francis Institute of Technology, Mumbai, India
e-mail: vb12781991@gmail.com

N. Ansari
e-mail: naz2ensfit@gmail.com

1 Introduction

In today's world, day-to-day human actions and interactions produce heaps of data containing hidden knowledge that can be processed and transformed into useful information. The activity of data mining (DM) is introduced to excerpt a more synthesized form of raw input. Improvements in the field of computer technology have led to a drastic increase in the usage of DM owing to its simplicity and affordability. Nevertheless, this has also had an influence on the obtainability of DM tools to work with; as such, the number of easily accessible tools has extended rapidly. Commonly used open-source DM tools and languages include Weka, Tanagra, scikit-learn, R, RapidMiner, Python, Orange, and KNIME. Therefore, it has become slightly challenging for an inexpert user to select the most apt software solution for his/her work. Under classification, there are various algorithms such as Naïve Bayes, decision tree, and ID3. This study focuses on comparing the performance of DM tools and programming languages like Weka, R, Orange, and Python by performing Naïve Bayes classification on student placement data.

2 Literature Review

A great deal of research has been performed for comparing open-source DM software tools based on various techniques and algorithms. A significant amount of the literature was reviewed for the same. Vrushali Mhetre et al. [1] performed prediction using Weka to identify slow, average, and fast learners where student details were taken from Sardar Patel Institute of Technology, Mumbai, India. They performed a comparison among various classification techniques such as Naïve Bayes, ZeroR, and random tree (RT) and concluded that RT performs best with 95.45% accuracy. Sharon Christa et al. [2] provided comparative analyses of Weka, RapidMiner, Tanagra, Orange, and KNIME and suggested integrating existing DM tools with agents to implement an effective data preprocessing architecture. Dr. Anil Sharma et al. [3] performed experimental analyses with two different datasets—i.e., iris dataset containing three different types of irises: *setosa*, *versicolor*, and *virginica*; *petal length*; and *sepal length* and unbalanced dataset that contains one class value in majority and other in minority—on Weka and compared Naïve Bayes, SMO, and classification tree. They concluded that the Naïve Bayes algorithm provides good results in a balanced dataset, and its accuracy decreases in an unbalanced dataset. They were of the opinion that SMO and classification tree provide better results in an unbalanced dataset. Jović et al. [4] described the qualities of six widely used free software tools for DM. They provided comparison of algorithms covering classification, regression, clustering, etc., and recommend that RapidMiner, R, Weka, and KNIME have preferred features for a copiously purposeful DM platform and can be used for several DM tasks. Paramjit Kaur et al. [5] provided comparative analyses of three decision tree algorithms (i.e., ID3, J48, and RT) to predict final-year student campus placement

results using Weka on a real dataset of the Department of Computer Engineering, Punjabi University, Patiala. Their study showed that the performance of RT is better than that of ID3 and J48 in terms of accuracy and build time. Jeevalatha et al. [6] used ID3, CHAID, and C4.5 to predict student placement from a dataset containing details of previous 2 years passed out pupils from placement section in Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, by using RapidMiner. Their results stated that ID3 was the best algorithm that provided 95.33% accuracy on collected placement data. Rakesh et al. [7] analyzed whether a student is placed or not using J48 through Weka. The study was performed on 116 MCA students of Ghaziabad's reputed engineering college. Namita et al. [8] suggested the use of ID3 to predict student placement and compared ID3 results with other algorithms like C4.5 and CHAID. They were of the opinion that ID3 is the best classifier that offers an accuracy of 95%. Ajay et al. [9] performed comparative analyses on Naïve Bayes, multilayer perceptron, and J48 using Weka to predict whether students can be placed after MCA course completion or not. They concluded that Naïve Bayes classification is the best algorithm with 86.15% accuracy and 0.28 average errors, which is the lowest as compared to other algorithms. Although the abovementioned systems are helpful in one way or the other, our study focuses on comparing four tools to perform DM, namely Weka, R, Orange, and Python. In this study, we have performed Naïve Bayes classification on student placement data to determine whether a student can be placed or not and compared the accuracy of the four tools. The results of our study prove to be promising for a user and will definitely assist him/her to select the best DM tool to analyze his/her data and come to a concrete conclusion.

3 System Model

Figure 1 shows the block diagram of the proposed system, wherein student placement data is used as an input that is preprocessed and Naïve Bayes classification is applied to perform prediction whether students are placed or not.

A method of determining a model that illustrates and extricates datasets or concepts for being able to use the model to foresee a group of objects whose class label is unfamiliar is known as classification. Classification is always predicted rules. The above system gives an overview of the DM structure of Naïve Bayes classification. The dataset consists of records of engineering students of CMPN and INFT branches,



Fig. 1 Block diagram of the proposed system

which acts as an input to different DM tools. Next, data preprocessing occurs. First, at the preprocessing stage, training data is given as input. The system thoroughly analyzes and understands the training dataset based on target attribute values. Second, classification is performed on the training dataset. Third, testing data is provided on which we actually desire to perform prediction. Note that testing and training datasets should be compatible w.r.t. the number of fields and attributes. Fourth, prediction is performed based on a selected attribute. Finally, different DM tools interpret other attributes in a different way and generate prediction results.

4 Naïve Bayes Classification Algorithm

It is an algorithm that utilizes Bayes' proposition to estimate how the extent of belief in a scheme alters with evidence. It utilizes Bayesian interpolation of likelihood where likelihoods represent the extent of belief. Bayesian learning functions in the following manner: Before any data has been perceived, the anticipation as to what factual association concerning those data might be can be stated in a probability distribution over the hypotheses that define this relationship. This division is referred to as *prior*. Once data has been publicized to the algorithm, reviewed discoveries are apprehended as a subsequent likelihood distribution. The assumptions that have seemed to be believable before, yet were discovered to be improper, will undergo a reduction in their likelihood. Conversely, suppositions that have accomplished to meet the anticipations will have their likelihood enlarged. The Naïve Bayes classifier is implemented by estimating a subsequent probability. Assuming a set of variables $X = \{x_1, x_2 \dots, x_n\}$, the algorithm tries to determine a subsequent probability of C_i amid a group of probable results $C = \{c_1, c_2 \dots, c_j\}$. Subsequent probability is calculated as follows:

$$p(C_i|X) = p(C_i) \prod_{k=1}^j p(x_k|C_i)$$

Using this modified Bayes' rule formula, a new case X is labeled with a class level C_i that attains the uppermost subsequent likelihood (Fig. 2).

5 Dataset Description

The dataset that is considered is of the year 2015 of B.E. students from CMPN and INFT background of St. Francis Institute of Technology (SFIT), Mumbai, India. A total of 200 records of students were considered. The objective behind using this dataset is to determine whether a student is placed or not {y/n} using four DM tools.

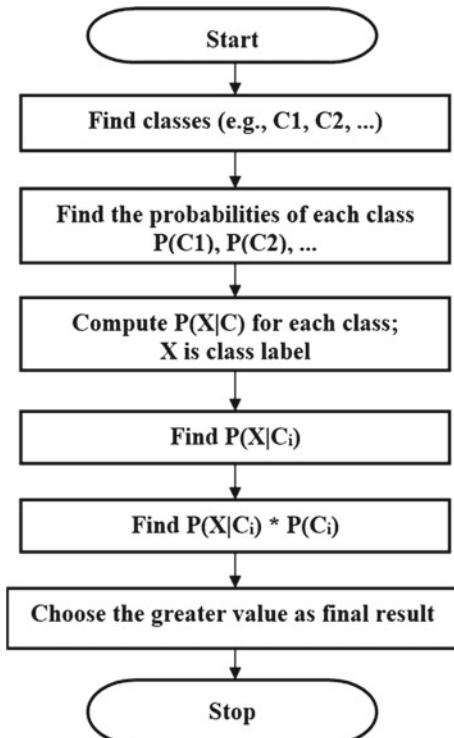


Fig. 2 Naïve Bayes classification algorithm

The accuracies of the tools were determined from the results achieved. Table 1 lists attributes, description, and possible values considered for student placement data.

6 Implementation in Weka

6.1 Training and Testing Datasets

The number of instances is 100 (50 from CMPN and 50 from INFT). In the dataset, some of the students were placed and some were not. The dataset has ten attributes, as mentioned in Table 1, out of which we selected *predict* attribute as the target attribute. In the target attribute, it shows if a student is placed or not (*yes/no*). The target attribute uses S.S.C., H.S.C., and EnggAgg as a reference to predict students' placement. The dataset was trained on the basis of the target attribute on four different DM tools. Next, the trained dataset was used on testing dataset to generate results on different tools. The *predict* attribute was inserted in the dataset to compare the

Table 1 Attributes, description, and possible values considered for student placement data

Attributes	Description	Possible values
Title	Gender of student	{Mr., Ms.}
First Name	Student's first name	
Middle Name	Student's middle name	
Last Name	Student's last name	
S.S.C.	Std. 10th percentile	
H.S.C.	Background of student	{H.S.C./DIPLOMA}
H.S.C./Diploma	Marks of student in H.S.C./diploma	
Branch	Branch of student	{CMPN, IT}
EnggAgg	Engineering aggregate of student	
Predict ^a	Whether a student is placed or not	{y/n}

^aPrediction is performed on **Predict** attribute

prediction accuracy of different tools. In actual, there were 100 instances considered to predict student placement.

Table 2 shows the output of testing dataset after training Weka for 100 instances. Weka checks for each and every instance, and by considering the target attribute for each instance, it makes prediction. Here, we can see that for testing 100 instances, Weka correctly classifies 68 instances, and 32 instances were incorrectly classified. Weka also shows the output in the form of a confusion matrix. The confusion matrix clearly tells the prediction made by Weka and which of the instances were incorrectly classified as correct instances and vice versa. For the above dataset, we can see that 58 instances were correctly classified as *yes* and 10 instances were correctly classified as *no*. Also, 8 correct instances were incorrectly classified as *no*, and 24 incorrect instances were incorrectly classified as *yes*. So, this helps to know the actual number of predictions that went wrong by Weka.

7 Implementation in Orange

7.1 Testing Dataset

See Table 3.

The above table shows a snippet of testing student placement data in Orange. Here, the target attribute selected is *predict*. So, majorly based on this attribute, further predictions are performed by the tool. The dataset comprises students from CMPN and INFT branches for more accurate results as their EnggAgg score varies based on lower grade marks.

Table 2 Prediction result of testing student placement data in Weka

Correctly classified instances	68	68%
Incorrectly classified instances	32	32%
Kappa statistic	0.1952	
Mean absolute error	0.3233	
Root-mean-squared error	0.4717	
Relative absolute error	82.1943%	
Root relative squared error	93.522%	
Total number of instances	100	

CONFUSION MATRIX

a	b	← Classified as
58	8	a = Yes
24	10	b = No

Table 3 Testing student placement data in Orange

Predict	S.S.C. (%)	H.S.C./Diploma	H.S.C. (%)	EnggAgg (%)
Y	90.92	H.S.C.	91.00	80.22
Y	93.80	H.S.C.	91.60	77.72
Y	91.23	Diploma	90.33	74.80

Table 4 Prediction result of testing student placement data in Orange

Naïve Bayes	Predict	S.S.C.	H.S.C./Diploma	H.S.C. (%)	EnggAgg (%)
0.05–0.95-Y	Y	90.92	H.S.C.	91.00	80.22
0.05–0.95-Y	Y	93.80	H.S.C.	91.60	77.72
0.05–0.95-Y	Y	91.23	Diploma	90.33	74.80

Table 4 shows the prediction result of testing student placement data in Orange. Herein, we see that the predictions made are displayed in the left corner with the attribute name as Naïve Bayes. In addition, the Naïve Bayes column shows the predictions made on the basis of probability. Note that the instance with more probability signifies correct prediction made by Orange. The red color bars below each probability show the distinct prediction. Therefore, we can observe and conclude that major parts of the instances were correctly predicted.

Table 5 Prediction result of testing student placement data in R

8 Implementation in R

8.1 Testing Dataset

Table 5 shows the prediction result of testing student placement data in R where the entire dataset is given to the tool as input. Next, we divide the dataset based on probability (we divided the dataset into half, i.e., 0.50–0.50 probability), which means that first 100 instances were selected for training the tool and the remaining 100 instances were selected for testing. It can be seen from `nrow->trainD` result and `nrow->testD` result. Next, we train the data in `trainD` via Naïve Bayes classification and store the result in model. Finally, we make use of predict command to compare the model and `testD` classes, and the output is displayed as `y` or `n`.

9 Implementation in Python

Python is executed on Jupyter Notebook. Preprocessed dataset is divided based on probability (we divided the dataset into half, i.e., 0.50–0.50 probability), which means that first 100 instances were selected for training the tool and the remaining 100 instances were selected for testing. Next, we created models *gclassifier* and *bclassifier* for Gaussian and Bernoulli's Naïve Bayes classification algorithms, respectively. These models are available in sklearn library. Later, the models are trained on the training dataset and used for prediction over the testing dataset using predict function. The results of Gaussian and Bernoulli's Naïve Bayes are stored in *gy_pred* and *by_pred*. Lastly, we display the number of correct predictions and also make use of

Table 6 Prediction result of testing student placement data in Python (Gaussian Naïve Bayes)

In [19]: from sklearn.metrics import accuracy_score print('Accuracy of "GaussianN8":', accuracy_score(y_test, gy_pred)) Accuracy of "GaussianN8": 0.67
--

Table 7 Prediction result of testing student placement data in Python (Bernoulli Naïve Bayes)

In [19]: from sklearn.metrics import accuracy_score print('Accuracy of "BernoulliN8":', accuracy_score(y_test, by_pred)) Accuracy of "BernoulliN8": 0.66
--

accuracy score function to display the percentage of correct prediction made by the trained model (Tables 6 and 7).

10 Experimental Results and Observations

Figure 3 shows the graphical representation of comparison among DM tools on student placement data. From the figure, we can infer that all the tools show similar predicted results. The more accurate is Weka. Weka has slightly higher accuracy as compared to that of other three tools with *68 correct instances* and *32 incorrect instances*. For second accurate results, there is a tie between Orange and Python (GNB) showing *67 correct instances* and *33 incorrect instances*. The least accuracy was found in R with *63 correct instances* and *37 incorrect instances*. We can see that almost every tool provides relatively same prediction, but if we consider for the most accurate result, then Weka offers the most accurate prediction of 68%. This means that out of 100 students, Weka correctly predicted the placement of 68 students based on their academic scores. Orange and Anaconda (P-GNB) provided an accuracy of 67%.

11 Conclusion and Future Scope

Naïve Bayes classification on different DM tools was effectively implemented on student placement data. From the above results, it is distinct that Weka provides more accurate results as compared to other tools. Apart from the accuracy, we observe a difference in the working styles and ease in access of the tools, wherein Weka and Orange involve basic selection and drag–drop functionality whereas R and Python involve coding commands. In addition to our study, we could consider a larger, more diverse dataset with many other algorithms and tools in order to get a broader perspective on the working of each. Moreover, the results of the algorithms can be used by Training and Placement Cell to recognize those set of students who are expected to confront difficulties during campus placement. In the future, we will also

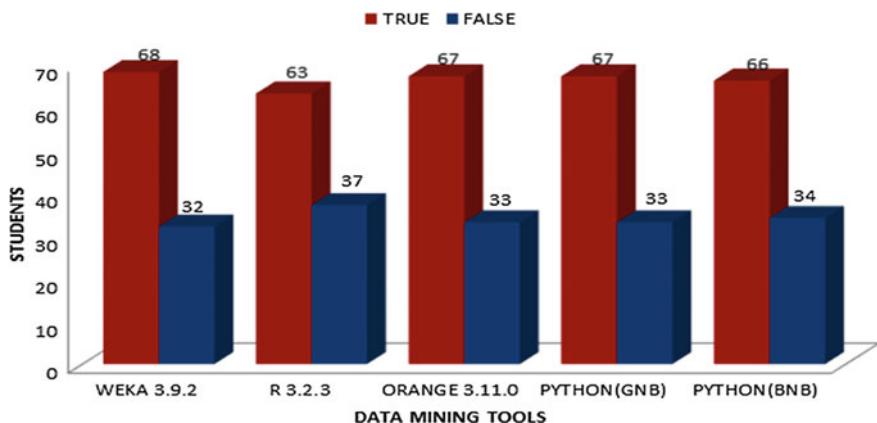


Fig. 3 Graphical representation of comparison among DM tools on student placement data

try to identify those students who need special attention and guidance to increase their likelihood of placements. We also plan to extend our dataset to EXTC students, and along with Naïve Bayes classification, we aim to perform clustering and association so that prediction can be performed much more accurately.

References

1. Vrushali, M., Mayura, N.: Classification-based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. In: Computing Methodologies and Communication (ICCMC), pp. 475–479 (2017)
2. Sharon, C., Madhuri, K.L., Suma, V.: A comparative analysis of data mining tools in agent-based systems. arXiv preprint [arXiv:1210.1040](https://arxiv.org/abs/1210.1040) (2012)
3. Anil, S., Balrajpreet, K.: A research review on comparative analysis of data mining tools, techniques and parameters. Int. J. Adv. Res. Comput. Sci. **8**(7) (2017)
4. Alan, J., Brkic, K., Bogunovic, N.: An overview of free software tools for general data mining. In: Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1112–1117 (2014)
5. Paramjit, K., Kanwalpreet, S.A.: Comparative analysis of decision tree algorithms for the student's placement prediction. Int. J. Adv. Res. Comput. Commun. Eng. **4**(6), 396–400 (2015)
6. Jeevalatha, T., Ananthi, N., Saravana Kumar, D.: Performance analysis of undergraduate students' placement selection using decision tree algorithms. Int. J. Comput. Appl. **108**(15), 27–31 (2014)
7. Rakesh Kumar, A., Dhannendra, B.: Placement prediction through data mining. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **4**(7), 447–451 (2014)
8. Namita, P., Deepali, K., Pratiksha, S., Kishori, B., Deepali, M.: Student placement prediction using ID3 algorithm. Int. J. Res. Appl. Sci. Eng. Technol. **3**(3), 81–84 (2015)
9. Ajay Kumar, P., Saurabh, P.: Classification model of prediction for placement of students. Int. J. Mod. Educ. Comput. Sci. **5**(11), 49–56 (2013)

Blockchain Technology: An Aid to the Governance of Smart Cities



Vishva Nitin Patel and Chhayaben Nitin Patel

Abstract Blockchain technology has spread its roots from the crypto-currency domain and found a way to handle the processes of data storage, data management, smart contracting, authentication, validation, trust-free transaction handling, and shared economy. In the era of smart cities, the shift from a system that depends upon conventional data management based on ledgers and government records to smarter, immutable and transparent blockchain mode of handling data and its transactions is inevitable. The success of ‘Bitcoin’ can be viewed as a proof of the success of blockchain technology, and thus, it should be seen as a breakthrough to solve various issues faced by digitization of transactions and crucial documents. The decentralized architecture of blockchain can support Internet of Things applications, governance of various smart businesses and smart cities beyond our imagination.

Keywords Blockchain technology · Decentralized architecture · Shared economy · Trust-free · Proof-of-work · Double-spending · Peer-to-peer network · Smart city · Smart contracts · Smart property · Hash function · Transparent governance

1 Introduction

The idea of a smart city has gained increasing popularity, and with the help of advances in Internet of Things, smart cities are becoming a reality. With the realization of the dream of a smart city, the question arises about how the governance of these

V. N. Patel (✉)

Department

of Information Technology, Madhuben and Bhanubhai Patel Women’s Institute of Engineering, Gujarat Technological University, Anand, Gujarat, India

e-mail: vishvapat96@gmail.com

C. N. Patel

Anand Institute of Information Science, Anand, Gujarat, India

e-mail: chhaya.n.patel@gmail.com

smart cities will be handled. As everything within the perimeter of the smart city will be digitized, the crucial task is to handle all the contracts, resource allocation, data management, and transactions within the city as well as in between various cities. While blockchain is usually linked with Bitcoin, it is also a breakthrough to enable a smarter model of governance. The blockchain technology is smart, distributed, shared, encrypted, and secure thus laying the core foundation for being trust-free, democratized, automatic, transparent, and private. It not only handles the transactions securely, it enables the transactions to be performed directly between the two parties involved, without the need of a ‘trusted’ intermediately thus ensuring a ‘trust-free’ mode of operation. It depicts the transparent link between data, its source, and all the manipulations performed on it along with the respective time stamp and the person who performed that manipulation. The immutable blocks of data enable us to apply blockchain model into the functioning of almost all the digital content held within the smart city.

2 Blockchain Technology: Beyond Bitcoin

While it is true that the success of Bitcoin has turned an increased interest in the blockchain technology, one cannot merely limit the application areas of this technology to digital currency and its transactions. The technology has to be viewed in the perspective of an information infrastructure to realize its full potential. Not only can blockchain support the economies lying within a smart city, but it can also significantly support the smart city’s governance and data management.

2.1 *Understanding the Block-Chain Model*

Basically, one may view ‘blockchain’ as a ledger or, more simply, a sequential database of transactions recorded by a network of computers. Essentially, in the Block-Chain Model, the ‘Block’ holds the data while the ‘Chain’ is the link that joins these blocks of data in a ‘time-stamped’ order. The flow of transactions or alterations made on the blocks of data can be precisely traced by traversing across the chain. This simplicity and transparency of data handling is one of the best features of blockchain. Blockchain is a peer-to-peer network, wherein all the members involved have a copy of each block, any amendments to these blocks, or the addition of a new block would require a majority in the consensus of all the members involved; moreover, when any change is made to an existing block, that particular copy of the block remains unaffected, and a new block containing the altered data along with the timestamp indicating the time of alteration is added to the sequence of blocks. Furthermore, the blockchain model can be categorized as Permissionless or Public Blockchain where no authorization is required to access or modify the

content held; the second categorization being Permissioned or Private Blockchain where only authorized members can access or modify the data [1].

2.2 The Security Aspect

The security concern can be well handled in the case of the blockchain model. Security comprises of confidentiality, integrity, and availability: Availability—only after authorization and only for the permitted manipulations; Confidentiality—private data is kept private; Integrity—no tampering is allowed [2]. In order to ensure the control over data access and performing any alterations within the network, the following mechanisms are employed:

Public–Private Key Cryptography. It enables the receiving party to determine whether the data or modification or transaction or message actually originated from the intended sender.

Cryptographic Hash Functions. Hash functions, hash tables, and special cryptographic techniques combine to form the resulting cryptographic hash function which can be directly applied to establish security and privacy protocols required for the functioning of blockchain technology [1].

As shown in Fig. 1, each block is assigned a unique digital signature, i.e., hash value along with the previous block's hash value. The first block has no entry made for the previous block's hash value, and hence, it is termed as the ‘genesis block.’

In case of a breach in security, and the invader tries to tamper with a particular block, the hash value for that block will change and the next block will no longer point to that tampered block, as shown in the Fig. 2. As for the blocks following the ‘tampered’ block, all of them will have incorrect hashes. Thus, the hash values ensure that the network gets notified in case of an outsider trying to manipulate the

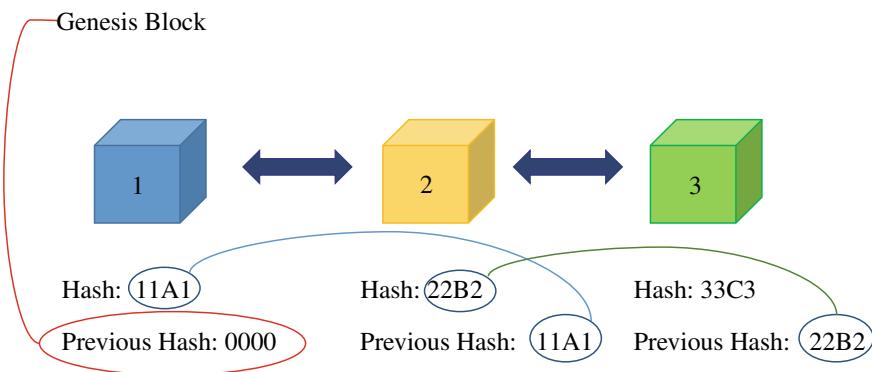


Fig. 1 Assigning hash values

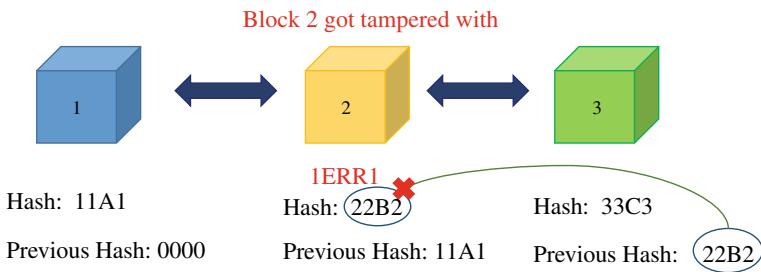


Fig. 2 Change in hash value when a block is tampered

blocks of data. But cryptographic hash functions alone cannot ensure the security of the blockchain; hence, an additional approach the ‘Proof-Of-Work’ was appointed.

Consensus through Proof-Of-Work. Since any transaction is validated only after attaining a majority in the consensus, it is a necessity that the result of the consensus be legitimate. Different voting mechanisms are employed to achieve consensus, the most widely used is Proof-of-Work, which depends on the amount of processing power invested in the network. If an attacker wants to manipulate any data within the blockchain, he has to commit a computational effort equivalent or greater than all the power spent from the reference point where he wishes to perform the alteration to the present time. In addition, he would have to achieve this faster than the current legitimate network processing of the new blockchain entries. Proof-of-work may be established based on solution to a computationally hard problem, a memory-intensive problem, or a problem that may require user interventions. To be practically useful for blockchain technology, such problems must be computationally challenging to solve, but efficient to verify a solution once obtained [1]. Although Proof-of-work ensures the security of the network, this mechanism is very costly in terms of computation and thus in electricity consumption as well.

3 Smart Cities and Shared Economy

Considering the success of Airbnb and Uber, shared economy seems to be the ‘next big shift.’ One can define shared economy as a peer-to-peer-based model of acquiring, providing, or sharing access to goods and services that are facilitated by a community network-based online platform. In the business perspective, shared economy represents the business approach wherein one does not need to ‘own’ a particular property, he can ‘share’ or ‘rent’ it as and when needed. As seen in the case of ‘Airbnb,’ where one can rent a place for a given time period and call it home. Because of the congestion, space constraints and the increased population tied to the urban living, smart cities need to be designed to be shared economies with consumption involving access to shared resources rather than asset ownership [2]. Blockchain enters the picture of shared economy since it can efficiently act as the immutable ledger for all the

shared transactions. Shared economy can act as the economic model which the huge population within a smart city can adopt to utilize the under-utilized assets while reducing the intermediate transaction costs. For the functioning of a blockchain-based shared economy in a smart city, smart databases collect information about the city's existing resources, smart control systems organize and schedule resource usage in a smart way, while people access these shared resources via smart interfaces [2]. The blockchain is a secure option that enables tracking down of the transactions made with respect to the shared resources, and these transactions cannot be tampered with. Blockchain ensures there is no double spending, and it is the best way to determine which resources are not utilized to their best capacity thus generating efficient plans for the full utilization of public and private resources available. Shared economy is based on the trust factor and in turn security and privacy, all of which can be handled seamlessly by blockchain technology, which eliminates the need for 'trust' and moves toward the 'trust-free' approach.

4 Going Trust-Free and Eliminating the Middlemen

Trust is one of the most crucial human factors in the functioning of any system that deals with monetary aspects and resource handling. When we consider the working of the blockchain technology, it takes away the role of 'trustworthy' third party and eliminates the need or question of 'trust' since it does not have any human involvement in its data transactions and it does not allow any outsider to manipulate the existing data. In case the blocks of data are tampered, the hash value and proof-of-work are effective measures to indicate the discrepancy. Blockchain has been termed as 'the trust machine' by The Economist, indicating that blockchain handles the trust issues between strangers involved in any sort of transaction [3]. The intermediator who simply adds an additional cost to the ongoing transaction is removed out of the picture completely. Privacy is also at stake once a middleman is involved along with the extra cost of having the middleman handle the transaction for us. In the scenario concerning smart cities and shared economy, the best option is to let the people handle their transactions by themselves, within the 'trust-free' environment. Furthermore, public and private keys are utilized to record the blockchain transactions thus ensuring anonymous and private transactions while enabling the 'verifier' to keep an eye on who exactly is involved in the transaction. This democratized transaction handling approach involves no middlemen and is completely secure and privacy oriented, thus giving a highly suitable option for handling the trust-free transactions a smart city truly needs.

5 Distributed Computing for Internet of Things and Distributed Data Management

Blockchain can act as the backbone for the working of IoT, since it can manage data from a variety of untrusted sources. The IoT consists of a gigantic network of Internet-enabled devices, of which some cannot be trusted and can even pose as a threat. The functioning of these network devices can rely on blockchain-based secure, private, and trust-free machine-machine and machine-human coordination. Blockchain technology can democratize and decentralize the functioning of smart cities that depends on the data provided by the devices, thus supporting distributed computing, which will ensure the future of the IoT [4]. Distributed computing significantly reduces the cost of processing billions of IoT transactions that take place daily as compared to the centralized data management architecture. Distributed data management is the pillar upon which the modern blockchain-based IoT stands, and the need for decentralization of authority and data is inevitable in the large-scale scenario. Since blockchain is a system based on peer-to-peer interaction, the entire data is spread across a web of computers thus making it extremely difficult to locate or modify data without the majority's consensus. Information can be published in encrypted format and can be distributed across thousands of different computers, making it impossible for any single node to modify [5].

6 Blockchain Application Areas that Support Smart Cities

The gradual shift from conventional networking, database model, and the dependency on Internet to the more secure, privacy-oriented blockchain model is inevitable once the features are compared. The blockchain-based computing of shared services facilitates the automation of business transactions and other services. The governance of smart cities can be completely digitized in a secure manner with the help of blockchain. Internet of Things (IoT) devices can achieve trust-free coordination and transactions, and contracts can be captured in computing codes to automatically perform all the obligatory actions that the parties involved have agreed upon in the given contract [2]. Following are some of the real-life application areas of blockchain technology along with their respective existing and successfully deployed applications which enable us to visualize the future of smart cities where blockchain is the core foundation.

6.1 Unique Digital Identity: Identification and Authentication

The existing ledger and record keeping systems render information about the residents of a particular city, but the security, privacy, and accessibility aspects can be properly handled by blockchain technology. Since everything is digitized within a smart city, the identity of its people is the first thing that needs to be digitized securely. A blockchain-based identity gives the residents complete control over who can access their personal information and how [3]. The need to carry identity proofs and authentications will be eliminated, at least inside the city. ‘ShoCard,’ ‘HYPR,’ ‘BlockVerify,’ ‘Onename,’ and ‘Guardtime’ are some of the prominent examples of companies dealing in the blockchain enabled identification and authentication procedure. ‘Civic’ aims at significantly reducing online identity frauds in credit reports by allowing its users to sign in and validate their personal identity-related information and lock that information. ‘UniquID Wallet’ has integrated fingerprint and other biometric identification mechanisms, and it can work on personal devices such as personal computers or smartphones and tablets.

6.2 Smart Contracts and Smart Properties

The automation of the entire marketplace and the shared economy can be made possible by the implementation of ‘smart contracts’ that are backed by blockchain technology. These smart contracts can deal with the transactions pertaining to any resource and that resource can be termed as a ‘smart property’ or Internet-based digital property. Several smart contracts can be combined and interact with one another to form decentralized organizations that operate according to specific rules and procedures defined by the collection of these smart contracts [5]. A new business operation model can be formed that has a decentralized notion of functioning, where the sellers and buyers can directly interact with one another. By providing a trust-free environment and coordination, blockchain enables new forms of collective actions within the predefined network, thus creating the potential to overcome the failures of the current model of governance. It can resolve a significant number of problems related to corruption and decisionmaking of various organizations. Trust does not lie solely with a selected few members of the given organization; instead, it is distributed within the network where the underlying protocol ensures security and auditability. All the operations can be scrutinized by the members of the peer-to-peer network. Furthermore, smart contracts bring a drastic increase in the speed with which contracts get executed and negotiated because here the smart contracts are not paper based and do not involve interpersonal involvement. Various open source projects such as ‘Ethereum’ [6], ‘Counterparty’ [7], and ‘Mastercoin’ [8] have been developed to create increasingly sophisticated smart contracts.

6.3 Smart Governance

When the transactions are directly recorded into the blockchain, we achieve a highly transparent and auditable trail of decisions made so far and that too at a reduced operational cost, thus leading to a successful model for smart governance [5]. Blockchain can bring more transparency to the working of the government thus enabling the residents to keep track of how the government is working toward providing better living standards for them. Frauds and scams can be significantly reduced since the blockchain model tracks all the manipulations made to the blocks of data. Blockchain can also effectively handle organizational governance, i.e., controlling the functioning of the organization by replicating the corporate model of governance and distributing the decision-making power amongst a well-defined group of parties via multiple signature technologies, hence preventing the execution of any action until all the parties involved agree to the transaction [5]. ‘BITNATION’ [9], the world’s first blockchain-based virtual nation, is a huge community where rights are distributed within the network and the identities are authentically established while being completely trust-free.

6.4 Authorship and Ownership

In a smart and digitized economy, the need to provide proof of ownership or authorship is mandatory. Smart cities can rely on the blockchain model to establish the direct link between the creator and the product, the owner and his property. Once this immutable link is generated, a concrete proof for further selling and buying of the property or product is in turn established. Various works of arts can be linked to their creators, and thus, no issues of stolen credit take place in the digital yet highly secure blockchain model. ‘Blockai,’ ‘Bitproof,’ ‘Verisart,’ ‘Stampery,’ ‘OriginalMy,’ ‘Monegraph,’ ‘Proof of Existence,’ ‘Crypto-Copyright,’ and many other companies are carving their way into authorship and ownership solutions through blockchain.

6.5 Energy and Resource Management

Smart cities are based on efficient planning and when it comes to the utilization of energy and other natural resources, it is necessary that the plan utilizes these resources efficiently, without any wastage. Transparency with respect to the amount of energy generated, where the energy was consumed, surplus energy and lack thereof can be clearly tracked by blockchain technology. ‘Energy Blockchain Labs’ [10] works on the energy revolution and is trying to develop energy-based Internet technologies built on top of blockchain. It covers energy production, consumption, trading, and management. ‘Grid Singularity’ and ‘TransActive Grid’ are similar examples.

6.6 E-Voting

To ensure complete security and untampered voting results, blockchain is an excellent option to look into. E-voting is not only limited to the working of the government, within the newly created decentralized business model, but this e-voting mechanism also ensures that the stakeholders are never jeopardized. In such a democratized environment of functioning, governments cannot tamper with the election results, and corporates cannot break the protocols of the organization [2]. ‘Follow My Vote’ is one such example of blockchain-based e-voting that offers a transparent and secure online voting system that ensures the accuracy and honesty of the election results. The company is developing a breakthrough allowing voters to independently audit the ballot box [3].

6.7 Job Market

The most important criterion in the job market is to check that the applicant actually is what his resume portrays. Thus, through the secure blockchain model, the verification process and all the identity-related concerns can be handled efficiently. Blockchain also enables the employer and the employee to communicate and finalize the contacts of work independently, without the need of an intermediary. ‘Appii’ uses blockchain to securely store and verify details of education, accreditations, awards, and employment history of the job seekers, and it allows the users to provide access to personal details to the extent they deem fit. ‘Coinality’ connects job seekers and employers where the salary is handled in the form of digital currencies and all the negotiations are handled independently between the two parties.

6.8 Media and Content Distribution

Not only can blockchain verify the authenticity of the content distributed, but it can also handle all the monetary transactions related to the renting or purchase of the content as well. No one outside of the well-defined peer-to-peer blockchain network can access the content being shared or tamper with it. Media of all sorts, from news and reports to social media, can become far more secure and free from discrepancies and fake content via the blockchain approach. ‘Bittunes’ handles music distribution by allocating the ultimate power directly in the hands of the artists and their fans. It uses Bitcoin as its primary currency, where the artists and music buyers earn Bitcoin automatically as part of the platform’s buying/distribution process. ‘PeerTracks’ and ‘Paperchain’ also follow similar approach.

7 Conclusion

After having a glance at the existing applications that have successfully employed blockchain technology as the backbone for their operation, it is hard to neglect the scope of this technology and its capacity to revolutionize the vision of an indeed ‘smarter’ city. The blockchain model brings simplicity and transparency, and it brings a democratized view into the governance system where all the stakeholders are aware of what exactly goes on within the system. It is a tool for bringing frauds and scams down to a significant low as it is a peer-to-peer network where misdoings hardly go unnoticed. The ‘trust-free’ smart contracts and transactions support the implementation of the shared economy architecture. Blockchain is basically a highly suitable data handling option for maintaining the governance and transaction-related operations of a digitized smart city.

References

1. Peters, G.W., Panayi, E.: Understanding Modern Banking Ledgers through Blockchain Technologies: Future of Transaction Processing and Smart Contracts on the Internet of Money
2. Sun, J., Yan, J., Zhang, K.Z.K.: Blockchain-based sharing services: What blockchain technology can contribute to smart cities. *Financ. Innov.* (2016)
3. <https://gomedici.com/21-areas-of-blockchain-application-beyond-financial-services/>. Last access date: 26/07/2018
4. Pureswara, V., Brody, P.: Device Democracy—Saving the Future of the Internet of Things. IBM Corporation (2015)
5. Wright, A., De Filippi, P.: Decentralized Blockchain Technology and the Rise of Lex Cryptographia
6. Ethereum: <http://www.ethereum.org>. Last access date: 26/07/2018
7. Counterparty: <http://counterparty.io/>. Last access date: 26/07/2018
8. Mastercoin: <http://www.mastercoin.org/>. Last access date: 26/07/2018
9. <https://www.cryptocoinsnews.com/bitcoin-governance-2-0-lets-blockchain/>. Last access date: 26/07/2018
10. <https://www.ibm.com/case-studies/energy-blockchain-labs-inc>. Last access date: 26/07/2018
11. <https://searchcio.techtarget.com/definition/sharing-economy>. Last access date: 26/07/2018
12. Atzori, M.: Blockchain Technology and Decentralized Governance: Is the State Still Necessary?

Machine Translation and Divergence Study for English–Maithili



Ritu Nidhi and Tanya Singh

Abstract In terms of language technology, Maithili is a resource-poor language. It is spoken in India and Nepal and is one of the 22 scheduled languages in India. Maithili has almost no language technology resource. English in India happens to be a dominant language in terms of content and usage. However, since more than 90% of Indians do not use English, a translation from English to Maithili becomes critical. An absence of basic tools in this language has affected resource creation of machine translation (MT). The present work discusses efforts for language technology resource (LTR) creation and divergence study for a statistical English–Maithili MT (EMMT) system. Creating any statistical MT (SMT) system requires sizeable parallel, aligned corpora for training and testing. Creating general-purpose source corpora for English language and creating translation equivalents with possible alignments require time and effort. The paper focuses on the data collection methods, cleaning, the size and structure of the text corpora, alignment and parallelization strategies, training, testing, and a study of divergence between the language pair.

Keywords Divergence · Machine translation · English–Maithili · Indian languages · MTHub

R. Nidhi (✉) · T. Singh
Amity University, Noida, UP, India
e-mail: jharitu@gmail.com

T. Singh
e-mail: tsingh2@amity.edu

1 Introduction

India is a multilingual country. According to the census of 2001, there are 1635 ratio-nalized mother tongues, 234 identifiable mother tongues, and 22 major languages.¹ MT systems are for automated translation from one language to another as human translation of any language is time-consuming and costly.

Maithili (ISO 639-3) is an Indo-Aryan language spoken in North Bihar and Nepal. As per 2001 census, there are 12,179,122 speakers in India. Including Nepal where Maithili is the second most spoken language, the total worldwide population is estimated to be 34 million.² (Figs. 1 and 2).

Maithili shows general features of an eastern Indo-Aryan language—rich inflection system, no gender agreement at the level of verbs, and flexible syntax leading to scrambling. However, it has its own unique features like the several degrees of

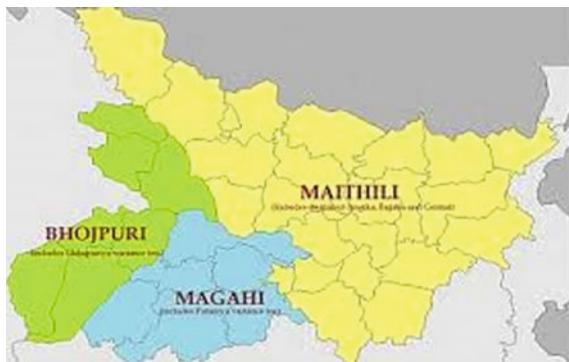


Fig. 1 Maithili speaking regions in Bihar (<http://blog.aaobihar.com/the-story-of-language-in-bihar/>)



Fig. 2 Maithili speaking regions in Nepal ([https://wikivisually.com/wiki/Mithila_\(region\)](https://wikivisually.com/wiki/Mithila_(region)))

¹“Census Data 2001: General Note”. *Census of India*. Retrieved 11 December 2014.

²<https://www.ethnologue.com/language/mai>.

honorifics at the level of pronouns and verbs. Though the syntax of Maithili exhibits a typical subject–object–verb (SOV) word order, increased scrambling often creates problems in automatic linguistic analysis. These unique features of Maithili are very distinct from English. Besides having a rigid syntax language with subject–verb–object (SVO) word order, English is weak in case inflections and morphology. Such and other distinctions between English and Maithili make the task of automatic translation more difficult.

Maithili is a low-resource language in terms of language technology development. Since translation technologies are not available for Maithili, the community's access to Internet where most of the content is in English is greatly constrained. India has 22 scheduled languages with English (spoken by less than 10% of Indian speakers) as a dominant language for governance and education. MT is therefore an important area of technology development for Indian languages. Even though English belongs to the Indo-European family of languages and shares genealogical connections with most Indian languages, yet it has remarkable differences with Indian languages, which any MT system has to consider for generating comprehensible outputs. Divergence is a phenomenon of linguistic incompatibility between a pair of languages that a translation process has to account for. However, development of a robust MT system including divergence-handling module for quality output has been a challenging task due to the fact that all possible cases of divergences between a pair of languages have not been identified and handled.

2 Resource and Method of Data Collection

Maithili suffers from a general lack of LTR which makes it difficult to create any tool for language processing or to collect more data. For training a SMT system, a basic dataset of 10,000 parallel sentences have been collected from various domains such as politics, sports, movie, mythology, cuisine, and tourism from sources like textbooks, newspapers, magazines, blog, and Web site (Fig. 3).

First, a generalized domain source language text for Indian English was collected from newspapers. Next, monolingual Maithili data was collected for training the translation model in target language syntax. Two native speakers including the author of this paper then manually translated the English data into Maithili. A third native speaker then checked the quality of the translation for accuracy. Popular alignment tool Giza++ was used for the alignment of the parallel data with poor results. The data were finally aligned and parallelized manually. After due checking and error corrections, the gold dataset was finalized. A third dataset of 10,000 sentences for the target language was also created as required by the MTHub³ platform (Table 1).

³<https://hub.microsofttranslator.com/Help/Index>.

Fig. 3 LTR for EMMT—data sources

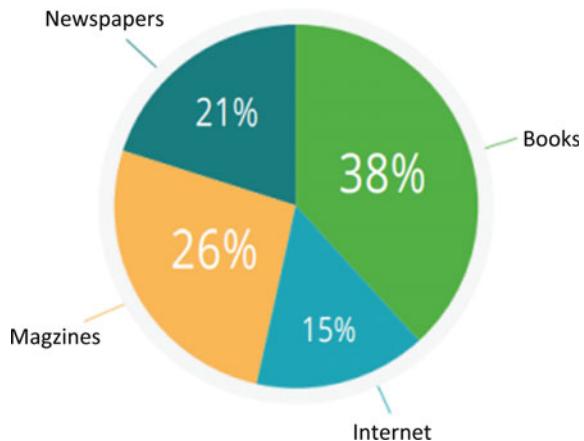


Table 1 Data size

English monolingual	Maithili monolingual	EM parallel
10,529	10,401	10,147

3 Challenges in Data Collection and Cleaning

Most Maithili speakers are bilingual with Hindi as the other language and there is a tendency to mix Hindi and other regional languages like Bhojpuri with occasional English. Therefore, it becomes necessary to clean the raw data before using it. Since it is extremely difficult to separate Hindi or Bhojpuri words/sentences from Maithili automatically as these languages share the script and many words/features, the task of cleaning the target language data was done manually. For removing English words, automatic script detection and cleaning program was used. In cases where English word was used deliberately as a case of code mixing and the automated process removed it, we had to replace it in the Devanagari script or with an equivalent Maithili word (Table 2).

Table 2 Mixing in the Maithili monolingual data and cleaning strategy

Language mixed	% mixed	Cleaning strategy
English	5	Removed automatically, added necessary data like email ids, URLs, etc
Hindi	15	Replaced with Maithili equivalents if not common
Bhojpuri	10	Replaced with Maithili equivalents if not common

Since there is no existing English–Maithili parallel data, we had to either translate English sentences manually as mentioned above or to auto-translate English to Hindi first and then translate Hindi to Maithili to save time. A rule-based Hindi–Maithili translator was created to do a rough translation of Hindi to Maithili which was later corrected to save time. The Hindi–Maithili rule-based translator works on a 200 handcrafted morphology rules and a transfer lexicon. Samples of some such rules are given below

#ती है=>त अछि	\$ही=>हे	#हुए=>ल	#ओं=>ा सभ
#आ =>लक	#आ रहा=>ायरहल	#ता है=>यछे	#एगा=>यत

4 Training the Microsoft Translation Hub (MTHub)

There are several platforms to train the system: MTHub and Moses are few of them. MTHub has been developed by Microsoft Translator to build translation system for translating between two languages using one's preferred terminology and style. MTHub was preferred over other competing platforms as it offers a good interface, linking the system to the community, evaluation, and deployment all in one place. The first training was done with 5 k data of the target language and 10 k parallel EM sentence-aligned data. The evaluation was done with the bilingual evaluation underway (BLEU) score.

Details of the training are given below (Table 3):

Total extracted sentence count: 31,077

Total aligned sentence count: 8027

Total used sentence count: 28,772

BLEU Score: 10.79

Table 3 Training data

Name	Type	Extracted sentences	Aligned sentences	Used sentences
MONOmai.txt	Monolingual	10,529	—	10,506
PARALLEL.txt	Parallel	10,147/10,401	8027	7880
PARALLELmai.txt	Monolingual	10,401	—	10,386

5 Divergence Study

Divergence occurs due to linguistic and cultural distances between languages. More the divergence, more skewed will be the results of the MT system. Therefore, divergence detection and handling as a pre/postprocess is critical for obtaining quality outputs. In a typical SMT system, the parallel text corpora of both source and target language should be significantly large to get better results. Dorr et al [1] mentioned, “Divergence occurs when structurally similar source language sentences do not translate properly into the target language.” A good study of divergence patterns in an SL-TL pair requires significantly large training corpora in a generalized domain. Mall and Jaiswal et al [2] have remarked, “Corpora are a body of language, which is an inevitable resource in natural language computing.” Nair and Davis Peter [3] have described, “Several MT systems that have been developed for Indian languages suffer from a paucity of dictionaries, rule-based, translation methodologies, and therefore require significant improvements to be usable.”

An analysis of the EMMT outputs, it was found that the causes of errors are varied and only some of them were actually due to divergent structures.

- Problem in transliteration, poor named entity (NER) recognition
- Difficulty in understanding metaphorical usage of verbs
- Poor verb dictionary due to insufficient training data
- Error in reference data
- Weak in reflexive recognition, **word order mismatch**, and **conflational divergence**

India came close to being ousted in the 1983 World Cup.

Ref: १९८३ विश्व कप मे भारत बाहर होबए के कगार पर आबि गेल छल।

MT: विश्व कप मे भारत close to being ousted आबि गेल छल।

- **Lexical divergence and word order mismatch**

Hazare continued his fast inside the jail.

Ref: हजारे जेल मे अपन भूख-हडताल जारी रखलखिन।

MT: हजारे जेल मे जारी भूख-हडताल।

- Failure in relative clause identification, **conflational divergence due to insufficient training data**

My vision is blurred both when I am looking far off and when I am looking close up.

Ref: हमर नजरि डिलमिलाएल रहैत अछि दुनु जखन हम दूर देखी या पास देखी।

MT: हमर नजरि blurred both when I am looking far off and when I am looking close up.

The translation outputs show the following categories of errors (a) NER failure (b) weak reflexives recognition (c) weak relative clause identification (d) word order mismatch (e) weak transliteration (f) conflational divergence and (g) lexical divergence. While solving cases a, b, c, and d will require specific re-trainings with targeted data in these domains or hybridizing the system with suitable rule bases, the errors due to divergence noted in f and g can be handled with strategies mentioned in the section below.

Another general pattern that emerges is that of insufficient training data being used in training the system. Many of clear cases of E-M divergence could not be detected due to a smaller than desired training set.

6 Handling Divergences

Dorr et al [1, 4] have described interlingual representation techniques (an extension of LCS of Jackendoff and others) as a solution for resolving MT divergences. Gupta et al [5] and Sinha et al [6] have remarked that the classifications of translation divergences as proposed by Dorr [4] is not sufficient to capture translation divergences for MT between English and Hindi bidirectionally. They have noticed, apart from Dorr's classification, additional list of translation divergences. Mishra et al [7] discussed English–Sanskrit translation divergences from the perspective of EBMT. Shukla et al [8] have defined seven cases of divergences between Sanskrit and Hindi at *vibhakti* level. Following a slightly different approach than Dorr's classification, Dave [9] has adopted interlingual approach to tackle the divergence issue between English and Hindi.

A statistical MT system for the English–Maithili language pair will have to prepare additional sets of data to re-train the MT model to take care of the above six cases of lexical semantic divergences. The process will be as follows:

- Training the MT model with English–Maithili parallel corpora aligned at sentence level.

An existing algorithm like the Microsoft Translation Hub (MTHub) can be used for carrying out model development. The parallel corpora will be prepared in sets of 10 K as per the requirements of MTHub. The training could be carried out for a specific MT project to develop a basic system

- Testing the model on new English sentences

The system so developed is tested on new data and results are evaluated automatically as BLEU score and also manually by judging the output under different parameters

- Error analysis to list errors and divergence types

The errors are analyzed against the gold data and comments with respect to one of the divergence types are added against each error

- Preparing additional corpora for retraining the model to minimize divergence errors

Additional training data are prepared for each type of error and a new system is trained with this data mixed in the new training set

7 Conclusion and Future Work

The paper discusses the resource creation efforts and divergence study for a statistical EMMT. Maithili is an under-resourced language and lacks basic tools. Creating a MT system will help data creation in the language and lead to creation of more tools. The training of the MTHub platform and study of errors have resulted in some divergences between the language pair, and additional data need to be created to handle them for improving the quality of MT outputs. Many of the errors were due to insufficient size of the training corpora, which needs to be substantially large to get comprehensible outputs. For a deployable EMMT system with good results, the size of the corpora should be increased and the system should be re-trained, tested, and evaluated several times. Preparing adequate additional parallel sentence-aligned corpora and training the EMMT system will be the next work and focusing on other kinds of divergences mostly syntactic will also be the subject of our subsequent papers.

Acknowledgements The lead author of this work is the native speaker of Maithili, which is the language of study in this paper. All the data reported and analyzed in this paper is her own and she gives her informed consent for this study.

References

1. Dorr, B.J., Jordan, P.W., Benoit, J.W.: A survey of current paradigms in machine translation. *Artif. Intell.* **63**(1–2), 429–492 (1993)
2. Mall, S., Jaiswal, U.C.: Developing a system for machine translation from Hindi language to English language. In: 4th International Conference on Computer and Communication Technology (ICCCT), IEEE (2013)
3. Nair, L.R., David Peter, S.: Machine translation systems for Indian languages. *Int. J. Comput. Appl.* **39**(1) (2012)
4. Dorr, B.J.: Machine translation divergences: a formal description and proposed solution. *J. Comput. Linguist.* **20**(4), 597–633 (1994)
5. Gupta, D., Chatterjee, N.: Divergence in English to Hindi translation: some studies. *Int. J. Transl.* **15**, 5–24 (2003)
6. Sinha, R.M.K., Thakur, A.: ●: In: Proceedings of EAMT, pp. 245–254 (2005)
7. Mishra, V., Mishra, R.B.: Study of example based English to Sanskrit machine translation. *J. Res. Dev. Comput. Sci. Eng.* **37**, 1–12 (2008)
8. Shukla, P., Shukla, D., Kulkarni, A.P.: Vibhakti divergence between Sanskrit and Hindi. In: Jha, G.N. (ed.) Sanskrit Computational Linguistics, LNCS 6465, pp. 198–208 (2010)
9. Dave, S., Parikh, J., Bhattacharya, P.: Interlingua based English-Hindi machine translation and language divergence. *J. Mach. Transl.* **16**(4), 251–304 (2001)

Detection of Liaison Between Health Outcome and Web Search



Sandip Roy and Amit Saha

Abstract Today Internet search has become a ubiquitous platform where people can assess their self-diagnosis of their health-related issues. Google Trends facilitates researchers to get this information related to their search behaviors toward health-related phenomena. These search-related topics and queries help to common people become a self-dependent which seems to be difficult to consult with their family, friends, or even a medical professional. Here we examined how the Google Trends search utility impacted to the leading cause of death in the USA. It is reported that around 75% of total all deaths in the USA are due to ten leading cause of death. We analyze search interest data using Google Trends API, and it is clearly shown that people are too aware on cancer-related issues followed by stroke and diabetes, where 22.5, 5.1, and 2.9 % of total deaths caused due to these respective diseases. This research analysis is demonstrated to understand how Google Trends data deeply reflected with the causes of death of millions of Americans.

Keywords Cancer · Diabetes · Google Trends · Internet · Stroke

1 Introduction

A recent trend demonstrates that around 72% of Americans have used Internet to diagnose their ailments rather than consult with the professionals [1]. Approximately 35% go for self-diagnosis and remaining take an appointment with medical professionals [2]. It has also reported by Google that one percent site's searches are connected to medical symptoms.

S. Roy (✉)
University of Kalyani, Kalyani, West Bengal, India
e-mail: sandip@klyuniv.ac.in

A. Saha
Wake Forest Baptist Medical Center, Winston-Salem, NC, USA
e-mail: aksaha@wakehealth.edu

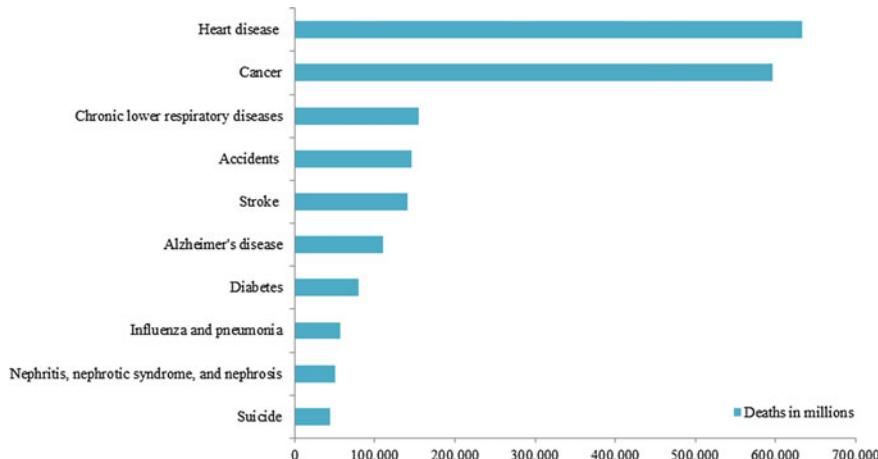


Fig. 1 Top ten leading causes of death in the USA

According to the WHO's statistics for ten causes of death globally in 2015, ischemic heart disease and stroke are the two biggest killer and around 15 million combined deaths are caused due to these reasons. Approximately 3.2 million people are died by both chronic obstructive pulmonary disease (COPD) and lower respiratory inspections. About 1.7 million deaths have happened for lung cancer including trachea and bronchus. Similarly, diabetes and Alzheimer's killed around 1.6 million people in 2015. Approximately 1.4 million deaths are caused by diarrheal, tuberculosis, and road accident [3].

In the USA, the picture is almost similar; here mostly four types of diseases are attributed around the people. They are mainly heart-related diseases, cancer, dementia, and diabetes; others are few from influenza, pneumonia, and kidney-related issues. Figure 1 illustrates the ten leading causes of death in the USA [4].

In this research, we use correlation statistical analysis to discover whether there is a relation between number of deaths from cancer, stroke, and diabetes in the USA and search-related queries on Google done by the Americans. In Fig. 2, it has also noted that in recent year, cancer, stroke, and diabetes are the most common causes of consciousness of the people. Due to this reason, we have concentrated on population-level real-time Internet search behavior which may be reflected with mortality rate in the USA.

2 Methods

Our research methodology is organized by the following steps below: (a) Data Sources (b) Statistical Analysis, (c) Discussion of Results.

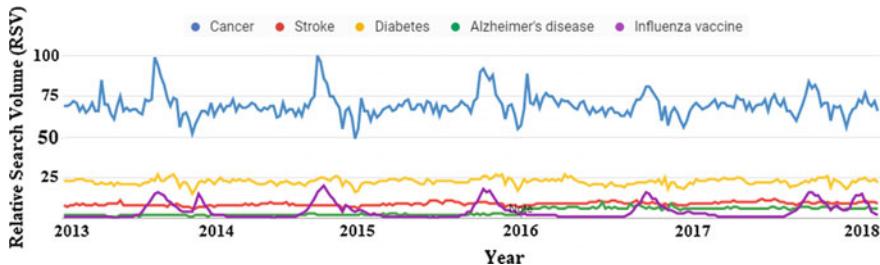


Fig. 2 Five years trend analysis of different leading causes of death in the USA

2.1 Data Sources

This research involved publicly available, free, online data source, where both Google Search interest data and age-adjusted mortality rates data for leading causes of death in the USA are listed from the year 2004–2017 and 1999–2015, respectively [5–7]. The common window in the data sources is 2004–2015, close to 11 years. In our experiment, we have only focused our result analysis on previous six years of the current decade which began on January 1, 2010.

2.2 Statistical Analysis

Globally, stroke is the second-leading cause of death behind ischemic heart disease and fifth-ranked cause of death behind heart disease, cancer, chronic lower respiratory disease, and accidental death in the United States. Each year, about 795,000 strokes occur; every forty seconds, someone has a stroke, and every four minutes, someone dies due to stroke in the USA [8].

In the USA, cancer is the second-leading cause of death and there are five common cancers like breast, lung, colon, prostate, and uterine. But globally, it is the fifth-ranked cause of disease [9]. Approximately, 595,930 deaths occur due to cancer in the USA and 609,640 cancer deaths are projected to occur in the year 2018 [10]. Diabetes, normally Diabetes Mellitus, is a major possibility for cardiovascular disease essentially coronary heart disease, stroke, and heart failure which are the most risk of death in the world. It was reported by the statistical fact sheet from 2011 to 2012; around 85.3% of children and 56.5% of adults had arrested due to this disease [8]. In 2014, the cause of mortality for 76,488 people in the USA due to diabetes and 245,016 deaths with diabetes as the primary or secondary cause of death [11]. In 2014–2015, globally, it was the sixth-ranked cause of mortality and seventh-ranked cause of mortality in the USA [3, 4]. Therefore, stroke, cancer, and diabetes are the most common cause of mortality in the USA. Due to this reason, people are more

aware and try to search symptom, prevention, and different health-care-related issues using different search engines like Bing, Google, and Yahoo.

In our research, our primary objective is to find out any relation between the mortality rates and searching frequency of people on these search engines in the USA. Here we have used Google Trends which is a free, open, Internet-based application that can permit the relative search frequency of different search topics or keywords that can also be compared each other over a specific time period [12]. It also gets the geographic location of the place from where the search terms or keywords used on the search engine. It can refer this search volume as a relative search volume (RSV) that value in between 0 and 100 where higher value of RSV represents most searched term and lower value of RSV mentions less searched term [9]. We have used to compare RSV of these searched terms, e.g., cancer, stroke, and diabetes variation over time according to nine divisions which are used for classification system and defined by the United Census Bureau [13, 14].

2.3 Discussion of Results

We examined nine regional divisions (like East South Central, West South Central, Pacific, Mountain, New England, South Atlantic, East North Central, West North Central, and Middle Atlantic) aggregate cancer, stroke, and diabetes death incidences using Centers for Disease Control and Prevention data and compare it with RSV on Google search engine from 2010 to 2015 [5–7].

Using the Pearson correlation [15], in Fig. 3, shows that regional division-level total RSV was notably correlated to cancer death incidence for year 2010 ($r = 0.75$, $p = 0.021$), 2011 ($r = 0.78$, $p = 0.012$), 2012 ($r = 0.79$, $p = 0.011$), 2013 ($r = 0.8$, $p = 0.0092$), 2014 ($r = 0.77$, $p = 0.015$), and 2015 ($r = 0.79$, $p = 0.011$).

In Fig. 4, regional division-level total RSV was also correlated to stroke death incidence for year 2010 ($r = 0.78$, $p = 0.012$), 2011 ($r = 0.8$, $p = 0.01$), 2012 ($r = 0.81$, $p = 0.0085$), 2013 ($r = 0.8$, $p = 0.0099$), 2014 ($r = 0.82$, $p = 0.0064$), and 2015 ($r = 0.82$, $p = 0.0068$).

Similarly, in Fig. 5, correlation between the RS and diabetes death incidence for year 2010 ($r = 0.76$, $p = 0.017$), 2011 ($r = 0.76$, $p = 0.018$), 2012 ($r = 0.74$, $p = 0.023$), 2013 ($r = 0.76$, $p = 0.018$), 2014 ($r = 0.75$, $p = 0.019$), and 2015 ($r = 0.76$, $p = 0.017$).

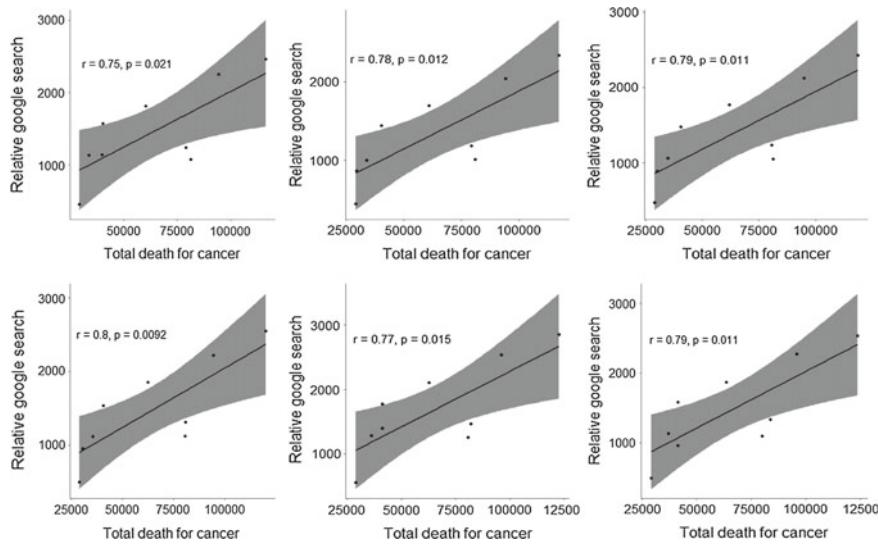


Fig. 3 Total death incidences versus relative Google search for cancer 2010–2015

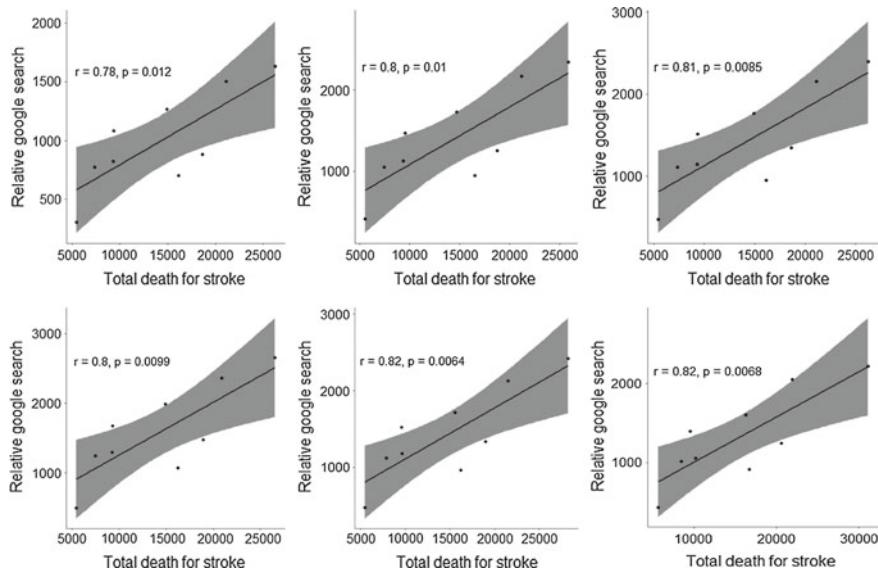


Fig. 4 Total death incidences versus relative Google search for stroke 2010–2015

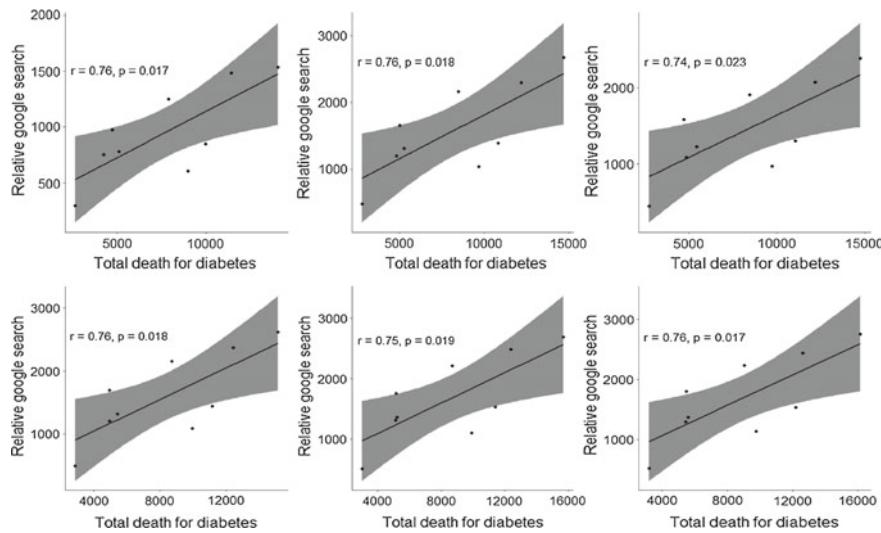


Fig. 5 Total death incidences versus relative Google search for diabetes 2010–2015

3 Conclusion

Google Trends data has extensive research directives, but our research has some limitations. In our research, we did not incorporate alternative search engine link Bing or Yahoo. Google Trends represented data only most common topic at the sub-region level or major metropolitan area. It was not suitable for rural areas and rare topic. Furthermore, Google Trends data was only accessible at the population level, and it cannot trace to a specific user. Additionally, Google Trends search was also reflected by public events such as awareness months and some fake news coverage by celebrity experiences which may be partially affected on our research.

References

1. Castillo, M.: More than one-third of U. S. adults use Internet to diagnose medical condition. CBS News (2013)
2. Aiken, M.: Googling our medical symptoms is making us sicker. QUARTZ (2016)
3. WHO.: The top 10 causes of death. Fact sheet (2016)
4. National Center for Health Statistics: <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>
5. Health searches by US Metropolitan area, 2005–2017 data from Google trends showing who searches for what and where. <https://www.kaggle.com/GoogleNewsLab/health-searches-us-county/data>
6. Searching For Health: <http://www.searching-for-health.com/>
7. NCHS—Leading Causes of Death: United States. <https://catalog.data.gov/dataset/age-adjusted-death-rates-for-the-top-10-leading-causes-of-death-united-states-2013>

8. American Heart Association Statistics Committee and Stroke Statistics Subcommittee: Heart Disease and Stroke Statistics—2017 Update A Report From the American Heart Association. **135**, 1–458 (2017)
9. Phillips, C.A., Leahy, A.B., Li, Y., Schapira, M.M., Bailey, L.C., Merchant, R.M.: Relationship between state-level google online search volume and cancer incidence in the United States: retrospective study. *J. Med. Internet Res.* **20**, 1–6 (2018)
10. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. *CA Cancer J. Clin.* **68**, 7–30 (2018)
11. Ali, M.K., Bullard, K.M., Saaddine, J.B., Cowie, C.C., Imperatore, G., Gregg, E.W.: Achievement of goals in U.S diabetes care, 1999–2010. *N. Engl. J. Med.* **368**, 1613–1624 (2013)
12. Google Inside Search: Insights into what the world is searching for—the new Google Trends. (2012)
13. Kittler, P.G., Sucher, K.P.: Most demographic and food consumption data are presented in this four-region format. *Food Culture. Cengage Learn.* **475** (2008)
14. Lewison, D.M.: Perhaps the most widely used regional classification system is one developed by the U.S. Census Bureau. *Retailing. Prentice Hall.* 384 (1997)
15. Shootman, M., Toor, A., Cavazos-Rehg, P., Jeffe, D.B., McQueen, A., Eberth, J., Davidson, N.O.: The utility of Google Trends data to examine interest in cancer screening. *BMJ Open* **5**, 1–8 (2015)

Wi-Fi-Based Portable Weather Station for Monitoring Temperature, Relative Humidity, Pressure, Precipitation, Wind Speed, and Direction



Indranil Sarkar, Bijoy Pal, Arnab Datta and Sandip Roy

Abstract This manuscript mainly focuses on the development of a model weather station to measure weather data: temperature, relative humidity, atmospheric pressure, wind direction, speed, and rainfall. This type of weather station has been designed to perform unmanned measurement of weather data. The measured data is wirelessly transmitted to the remote station for logging and displays the information to different smart gadgets. This wireless connectivity has been planned using Wi-Fi connections which establishes mesh network for reliable data communication. Furthermore, our consistent outputs do help the dwellers to take necessary precautions.

Keywords Cloud computing · Fog computing · Internet of things · Mobile weather station · Smart city

1 Introduction

Air quality monitoring and climate change have received much consciousness in the urban area. Urban dwellers always keep update about the latest weather conditions of their surroundings. Due to the availability of low-cost sensor platforms and the presence of high-speed reliable wireless data communication network, many devices are easily communicated through their measured data with each other and form mesh network [1, 2]. The average cost of sensors was \$1.30 in the year 2004, and it is

I. Sarkar
Elite Institute of Engineering and Management, Kolkata, India
e-mail: indra.nil2004@gmail.com

B. Pal · A. Datta · S. Roy (✉)
Brainware Group of Institutions, Kolkata, India
e-mail: sandiproy86@gmail.com

B. Pal
e-mail: bijoypal32@gmail.com

A. Datta
e-mail: dattaarnab2013@gmail.com

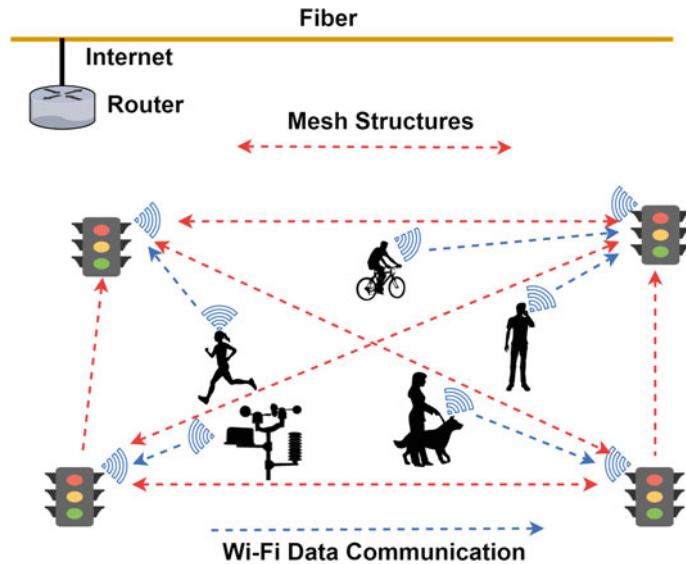


Fig. 1 Wi-Fi data communication for smart city

expected to reach \$0.92 in the year 2020 [3, 4]. In [5], Gartner forecasted that around 70% of the world's population is anticipated to live in cities by 2050. Due to the rapid urban growth in the next coming year, it will also make pressure on the existing infrastructure.

To assist the new demand on cities, Internet of things (IoT)-based system would reduce costs, improve services, and enhance communication and interaction [6, 7]. It is also very important that the best wireless connectivity technology for an IoT-based personal device to industrial instruments can be quite exacting (see Fig. 1) [8]. In [9], the authors discussed the various emerging wireless technologies for IoT.

Popular ISM bands like 433 MHz, 868 MHz, 915 MHz, 2.4 GHz, and 5 GHz are used in the wireless communication system. The unlicensed 2.4 GHz band became very popular, and it is used in all regions for connecting IoT-based devices [10].

Nowadays, air pollution is the most remarkable concern for public health. World Health Organization (WHO) suggested that each year around 5.5 million people die due to the pollution. So, many researchers work on real-time air pollution monitoring and forecasting system [11, 12]. It is also notable that IoT-based systems can also reduce the hardware cost into one-tenth of the previously used technologies. In [12], the authors focused on Zigbee-based real-time air pollution monitoring unit for smart city. Urban dwellers are not only aware of the air pollution monitoring unit but also have concern on the various weather attributes like temperature, relative humidity, pressure, precipitation, wind speed, and direction.

Therefore, in this manuscript, we have researched on portable weather monitoring station using Wi-Fi network technology.

Sections of the manuscript are as follows: Section 2 discusses the literature survey. Section 3 describes the experimental setup and result analysis. Section 4 deals with conclusions and examines the future scope of our research.

2 Literature Survey

Automatic weather station (AWS) is an unmanned system which saves human labor. A typical AWS consists of a rechargeable battery, data logger, meteorological sensors with attached solar panel or wind turbine, and the telemetry [13, 14].

Lage et al. described a mini weather station for crop protection [15]. Their system can also forecast the occurrence of Zonda winds before advance of several hours. This type of system was very much efficient for farmer. Saini et al. proposed the Arduino Uno board and Zigbee wireless technology supported weather station which can monitor various weather parameters and weather forecasting and alert the people from its catastrophic effect [16].

Thaker et al. proposed the ESP8266-based low-cost Wi-Fi-based system with Linux-based Web server application [17]. In [18], the authors described real-time weather monitoring system based on a mobile application. Jianyun et al. discussed IoT-based automatic weather station. The authors discussed the various features of the IoT-based automatic weather station like high-quality meteorological data, comprehensive monitoring information, reliable meteorological equipment, flexible scalability, standard interoperability, convenient deployment, and efficient maintenance and protection [19]. In our experiment, we have planned to organize an outdoor weather station for monitoring various atmospheric attributes and send the data to the remote cloud server.

3 Experimental Setup and Result Analysis

At first, we have organized our portable system on the outdoor platform unit as shown in Fig. 2a. In our system, we have used weather shield (see Fig. 2d) which was designed for SparkFun RedBoard and Arduino Uno, and also connected with weather meter [20, 21]. It is also noted that humidity/temperature sensor (HTU21D), barometric pressure (MP3115A2), tipping bucket rain gauge, cup anemometer, and wind vane are used in our proposed system which are illustrated in Fig. 2b, c [22]. The tipping bucket rain gauge, wind vane, and cup anemometer are connected with the weather shield by socket 6P4C [22]. We have used 9 V battery for our experiment as outdoor power supply unit.

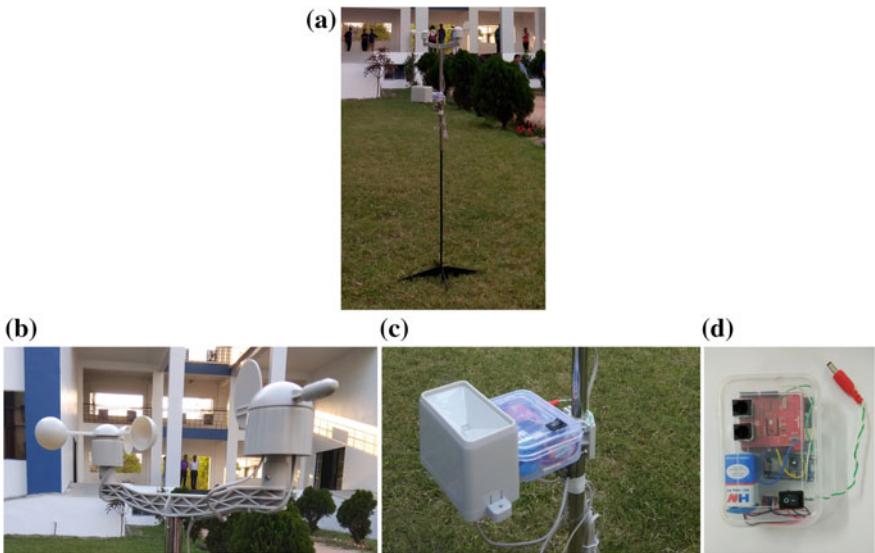


Fig. 2 **a** Mobile weather station, **b** wind speed and direction sensor, **c** rain gauge sensor, **d** weather shield

NodeMCU, an open-source IoT platform, is used for data communication. It can share the sample data with public channel Id #431128 of open-source IoT application and API named ThingSpeak.¹ Figure 3a–f illustrates the results of our experiment.

4 Conclusion and Future Scope

In this manuscript, we have designed an outdoor portable weather station. Our research objective is to not only sense data but also send the data to the remote cloud server using NodeMCU Wi-Fi module. Our research methodology will be more robust if we will include different forecasting techniques which can predict the results and minimize the error [23].

¹ Weather Station: <https://thingspeak.com/channels/431128>.

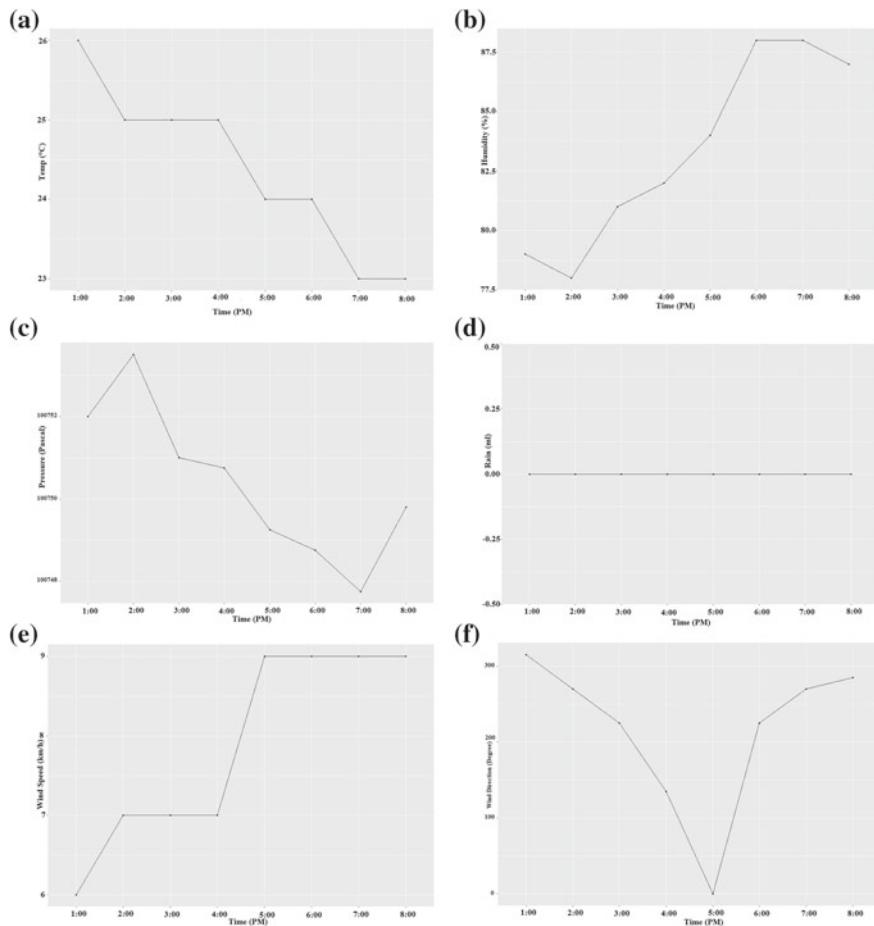


Fig. 3 a–f Graph of temperature, relative humidity, atmospheric pressure, rainfall, wind speed, and direction versus time: April 4, 2018

References

1. Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., Bartonova, A.: Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Env. Int.* **99**, 293–302 (2017)
2. Lopez, J.C.B., Villaruz, H.M.: Low-cost weather monitoring system with online logging and data visualization. In: 2015 International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, Cebu City, Philippines (2015)
3. Goldman Sachs, BI Intelligence Estimates: The average cost of IoT sensors is falling. ATLAS (2018)
4. Honrubia, M.: Industrial IoT is Booming Thanks to a Drop in Sensor Prices. ennomotive (2017)

5. State of the Market: Internet of Things 2016 Accelerating Innovation, Productivity and Value. Verizon.com (2016)
6. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: Internet of things for smart cities. *IEEE Internet Things* **1**, 22–32 (2014)
7. Kim, T-H., Ramos, C., Mohammed, S.: Smart City and IoT. *Future Gener. Comput. Syst.* **76**, 159–162 (2017)
8. Pau, G., Claudet, C., Dixian, Z., Collotta, M.: Next Generation Wireless Technologies for Internet of Things. *Sensors*, Editorial, pp. 1–5 (2018)
9. Elkhodr, M., Shahrestani, S., Cheung, H.: Emerging wireless technologies in the internet of things: a comparative study. *IJWMN* **8**, 67–82 (2016)
10. Lethaby, N.: Wireless connectivity for the Internet of Things: one size does not fit all. Texas Instruments Incorporated, pp. 1–16 (2017)
11. Xiaojun, C., Xianpeng, L., Peng, X.: IOT-based air pollution monitoring and forecasting system. In: 2015 International Conference on Computer and Computational Science, Noida, India (2015)
12. Roy, S., Bose, R., Sarddar, D.: Smart and healthy city protecting carcinogenic pollutants. *IJAES* **12**, 1661–1692 (2017)
13. King, J.: Automatic Weather Stations. Internet Archive 1 (2009)
14. Paul, D.: Watershed sensor network non-line-of-sight data telemetry system, Electronic Thesis or Dissertation, University of Cincinnati, pp. 1–63, (2007)
15. Lage, A., Correa, J. C.: Weather station with cellular communication network. In: 2015 XVI Workshop on Information Processing and Control, Cordoba, Argentina (2016)
16. Saini, H., Thakur A., Ahuja, S., Sabharwal, N., Kumar, N.: Arduino based automatic wireless weather station with remote graphical application and alerts. In: 2016 3rd International Conference on Signal Processing and Integrated Networks, Noida, India (2016)
17. Thaker, T.: ESP8266 based implementation of wireless sensor network with Linux based web-server. In: Symposium on Colossal Data Analysis and Networking, Indore, India (2016)
18. Munandar, A., Fakhrurroja, H., Rizqyawan, M.I., Pratama, R.P., Wibowo, J.W., Anto, I.A.F.: Design of real-time weather monitoring system based on mobile application using automatic weather station. In: 2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology, Jakarta, Indonesia (2017)
19. Jianyun, C., Yunfan, S., Chunyan, L: Research on application of automatic weather station based on internet of things. In: IOP Conference Series: Earth and Environmental Science, vol. 104, pp. 1–6 (2017)
20. Weather station wirelessly connected to Wunderground. SparkFun, 1 (2018)
21. Weather Sensor Assembly p/n 80422. Argent Data Systems, pp. 1–2 (2017)
22. Weather Meter Hookup Guide. SEN—08942 ROHS, 1 (2018)
23. Roy, S., Biswas, S. P., Mahata, S., Bose, R.: Time series forecasting using exponential smoothing to predict the major atmospheric pollutants. In: ICACCCN 2018, Greater Noida, India (2018)

Automated Socio-psychological Knowledge Behavior Classified in E-Commerce Applying Various Machine Learning Techniques



P. Vijayaragavan, R. Ponnusamy and M. Arrmuthan

Abstract Nowadays, the Web is a fundamental instrument of the search for a great many people. The Web returns a massive number of pages of result for every single search element. The Internet is an active medium for correspondence among personal computers and gets to on the Web records, yet it is not an instrument for finding or sorting out data. Tools like Web indexes help clients in finding data. The measure of information every day seeks on the network is vast, and the assignment of receiving intriguing and necessary outcomes rapidly turns out to be exceptionally taxing. The utilization of a programmed site folio classifier can improve the procedure through helping the Internet searcher in receiving relevant outcomes. The Web pages can exhibit unique and various data relying upon the attributes of its substance. The idea of Web content presents extra difficulties to site page arrangement when contrasted with conventional content characterization; yet, the interconnected concept of hypertext additionally includes and can help the process. The proposed system efficiently classifies the data applying various machine learning techniques like SVM and confusion matrix, K-mean clustering and aspect extraction. Machine learning procedures are of great significance, and it will be utilized to build the classifiers.

Keywords Support vector machine · Confusion matrix · Clustering · K-mean clustering · Text classification

P. Vijayaragavan (✉)

Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University, Kakinada 533003, AP, India
e-mail: vijayaragavanpm@yahoo.co.in

R. Ponnusamy

Department of Computer Science And Engineering,
CVR College of Engineering, Vastu Nagar, Mangalpalli(V), Ibrahimpatan(M),
Hyderabad 501510, Telangana, India
e-mail: r_ponnusamy@hotmail.com

M. Arrmuthan

Department of Information Technology, Perunthalaivar Kamarajar Institute of Engineering
and Technology (PKIET), Nedungudi, Karaikudi, India
e-mail: aramudhan1973@yahoo.com

1 Introduction

The perfection of data sources released on the Internet has made it exclusive for customers to utilize automatic categorizing in finding the preferred information resources. There is a need for creating shrewd structures for servers and clients that can make enough burrows for data. Web mining can be thoroughly portrayed as the disclosure and examination of significant data from the Internet by the retrieving of captivating and perhaps essential cases with specific data [1]. The issue concerning its versatility and future use does not relate to the page outlining, however, to the fitting affiliation and portrayal of its substance. This issue moreover impacts online business application, concerning the ways we find stock and ventures offered on the Web. Common looking engines discover Web pages by strategies for a lexical examination among sets of words and the hypertext's substance [2, 3]. The consumer socio-psycho knowledge behavior analysis in the online shopping became a contemporary research area for an increasing number of researchers to understand the unique nature of the customer behavior towards online shopping, their liking, disliking, and satisfaction level [4]. Semantic Web was announced in 2000 for this purpose, with its execution at first evaluated for quite a while. We display here the application and appraisal of SVM technique to see Web pages containing things accessible to be bought [5].

1.1 *Support Vector Machine*

Gathering, the buyer leads from customer educational accumulation. Decision line brought after with two edge lines with which data is close along the edge line is called reinforce vector [5]. Perplexity matrix-based reweighing can be utilized to accomplish certain sorts of changes in course of action while keeping up an equivalent level of precision.

1.2 *Confusion Matrix*

In the wake of arranging the information, apply disarray network, and this perplexity grid will distinguish not just positive and negative of all the likelihood, yet additionally will anticipate the great and terrible [6].

1.3 *K-mean Clustering Algorithm*

K-mean grouping is an unsupervised learning system and a non-hierarchical method-grouping of information sorted from the comparative sort or gathering.

Applying disarray framework printed data into information eluding separation measure will recognize the inside purpose of information, grouping of the information and in light of the new information be arriving. The middle information will vary at the current circumstance; specific E-business has been pushed so quickly that millions or trillions of clients are procuring things and tries online over the Web. The previous system did not group the information legitimately and successfully.

2 Proposed System

Arrange the information appropriately and adequately, and each and every content deliberately screens and analyzes it and while applying machine learning techniques classifies viable way [7, 8].

3 Implementation

Technique 1: Support Vector Model (SVM)

Classifying the online shopping consumer purchasing behavior using support vector machine technique [9] (Figs. 1 and 2).

SVM Model Algorithm

- Step-1: Initialize all data into an entire data set.
- Step-2: Use random forest algorithm to rank the features.
- Step-3: Evaluate aggregate voting to the ranked features.

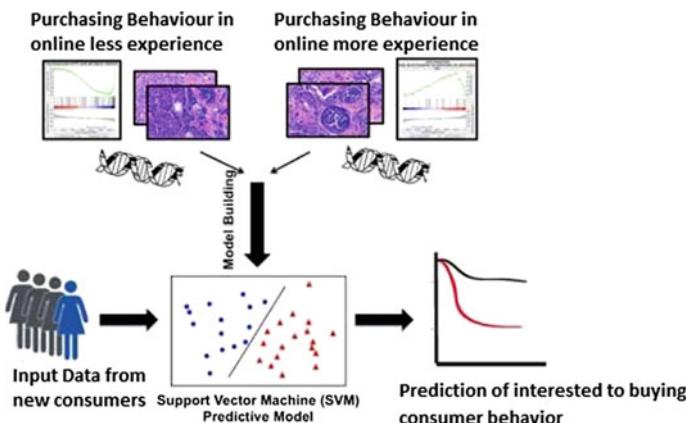


Fig. 1 Classify the data using SVM model

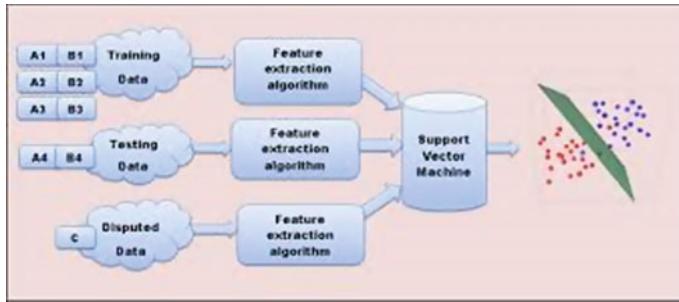


Fig. 2 How to classify the data from multi-source

Step-4: Update working data set by removing less important features, until the numbers of features are small.

Step-5: Apply SVM with RBF kernel into the reduced features.

$$\min \left(\frac{1}{2} w^T w + C \sum_{i=1}^m \xi_i \right) \quad (1)$$

Subject to

$$y_i (w^T x_i + b) \geq 1 - \xi_i, \quad i = 1, K, M \quad (2)$$

where w is weight vector and b is bias. The variable ξ is positive slack variable, which is necessary to allow misclassification.

Step-6: Consider parameter C seeks to penalize decision error when searching for the maximum marginal hyperplane.

$$y = f(x) = \text{sgn}(w^T \phi(x) + b) \quad (3)$$

where ϕ is a mapping function and is training sample.

Step-7: Choose the best parameter for creating the classification model.

SVM classifier is used to classify the consumer purchasing behavior on the online shopping consumer dataset and the results are shown in Figs. 3 and 4.

Technique 2: Confusion Matrixes A confusion matrix contains data about real and anticipated groupings done by an arrangement framework. Execution of such structures is ordinarily assessed utilizing the information in the matrix (Fig. 5). The accompanying table demonstrates the confusion matrix for a two-class classifier.

Fig. 3 Classify the data set decision line using SVM model

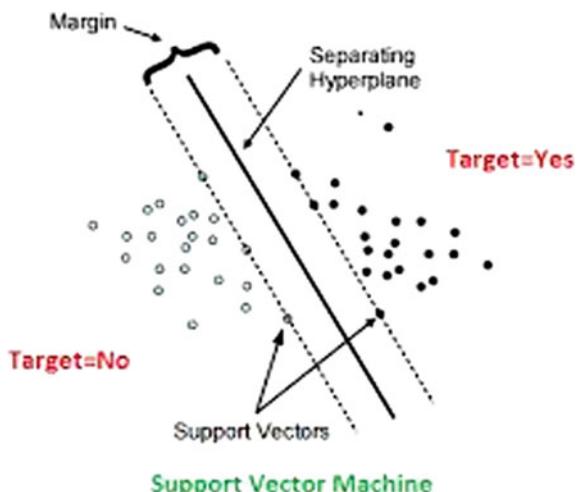
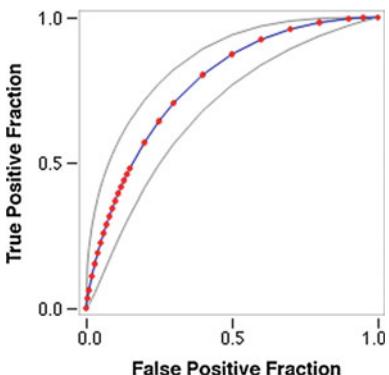


Fig. 4 ROC curve



Technique 3: K-mean Clustering

K-mean Clustering Algorithm (Fig. 6):

- **Input:** A set, V , consisting of n points and a parameter k .
- **Output:** A set X consisting of k points (cluster centers) that minimizes the squared error distortion $d(V,X)$ over all possible choices of X .

		Actual Result/Classification		
		Yes	No	
		tp(True Positive)	fp(False Positive)	Precision=tp/(tp+fp)
Predictive Result/Classification	Yes	fn(False Negative)	tn(True Negative)	Recall=tp/(tp+fn)
	No			
		Accuracy=tp+tn/tp+tn+fp+fn		F=2.precision.recall/precision+recall True Negative Rate=tn/(tn+fp)

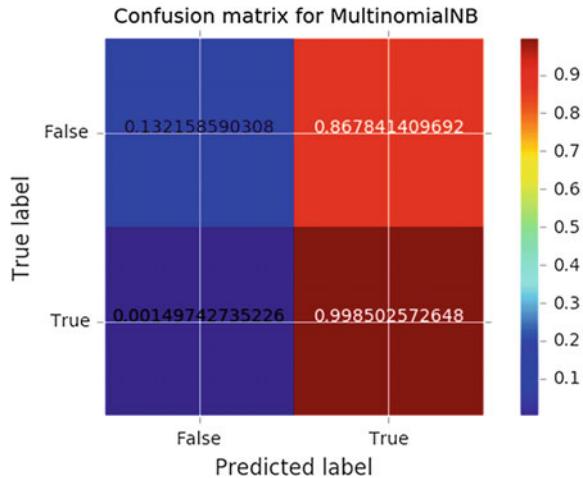
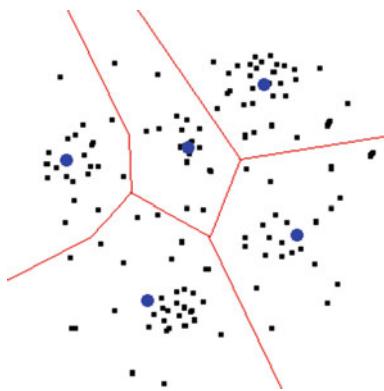


Fig. 5 Confusion matrix for Naive Bayes

Fig. 6 Output of K -mean clustering



- **K-mean Clustering**

1. Select a number (K) of cluster centers.
2. Assign every data point (e.g., gene) to its nearest cluster center.
3. Move each cluster center to the mean of its assigned data points (e.g., genes).
4. Repeat 2–3 until convergence.

Technique 4: Clustering and Aspect Extraction

The output of the classification algorithm is a labeled set of user reviews. The labeled positive reviews and the negative reviews are taken separately, and two clustering models are trained upon them individually. Each review is first predicted as either a good review or a bad review. With the predicted label, the words are checked for the distance between its position and the cluster centroid. The words which are closest to the centroid are the aspects of the products described in that particular review [10]. For a product, a complete word matrix of all reviews is created individually, and the less meaningful words are omitted. Then, the word's score is summed together to get a collective total score for each word in the matrix. Now, this word is checked for the distance between the centroid and its location using Euclidean distance. All the words which lie closest to the centroid are extracted from the matrix and displayed to the user as the aspects of the product which are described the most in the user reviews [1]. Some examples of the extracted aspects for the positive and negative reviews are given below:

Positive Review

I have been a k -espresso client for quite a while, and I particularly like Timothy's. I requested this seasoned espresso and the consistent decaf at a decent cost with the Amazon one-day transaction. All the k -container assortments are great, and your k -pot makes a crisp glass in 30 s. Amazon conveyed the espresso in only several days despite everything I'm getting a charge out of it. Much appreciated Amazon and Keurig when are you going to have another deal? I do not falter to buy at the standard cost since I truly appreciate the espresso; however, deals enable us to purchase more.

Extracted Aspects:

1. Transaction
2. Sec
3. Client
4. Appreciate
5. Truly.

Negative Review Nespresso makes great espresso and great machines. I changed over to a Nespresso machine 7 years prior and have never thought back. I spare a little fortune consistently by making my lattes at home. That being stated, the Nespresso case offers posted here are from an outsider who is putting an expansive

extra boundary on their cost. You can arrange similar items online from Nespresso for approximately 0.55 *for every* (a large portion of the cost here), and you can designate precisely what you need to purchase as opposed to taking a blend of things, some of which you dislike.

Extracted Aspects:

1. Nespresso
2. Boundary
3. Designate
4. Dislike
5. Fortune.

4 Structure of the Entire Process

The overall structure of the proposed work is shown in Fig. 7. Initial process gets the input of online shopping consumer purchasing behavior data set. Second process collection utilizing support vector machine technique classifies the data of the interest to obtaining purchaser and not interest buying consumer. Third process confusion matrix contains data about real and anticipated groupings done by an arrangement framework. Fourth process is a non-hierarchical method-grouping of data arranged from the relative sort or gathering. The fifth process is that each audit is first predictable as either a decent survey or an awful audit; with the expected mark, the words are checked for the separation between its position and the group centroid. The words which are nearest to the centroid are the parts of the items correspond to in that specific review. Final result is represented as statistical manner.

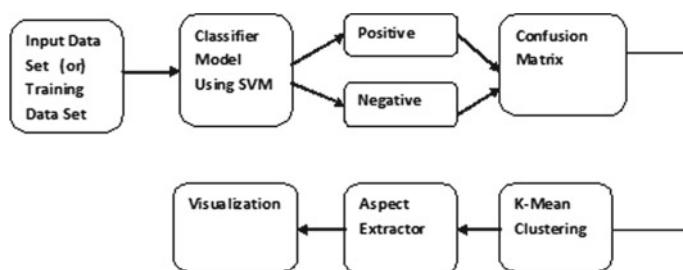


Fig. 7 Structure of the entire process

5 Conclusion

Purchaser audits characterize, or ordering in a compelling way to apply different machine learning procedures helps to recognize the shopper surveys and gives the ideal answer for the merchant and additionally the purchaser for increment of the item offers. It is useful to distinguish the shopper psychology in view of that vendor can offer the item since including psychology-based business future going to occur. The adequacy of the classifier depends on the personality of the optimistic and pessimistic cases chosen to create the preparation of data. One more important point tended to in this work comprises on the adjustment of the errands engaged with a request to dissect the space. The proposed work on online shopping can be employed in various real time applications.

References

1. Fiol-Roig, G., Mir-Juli, M., Herraiz, E.: Data mining techniques for web page classification. *Highlights in Practical Applications of Agents and Multiagent Systems*. AINSC **89**, 61–68 (2011)
2. Markkandeyan, S., Indra Devi, M.: Efficient machine learning technique for web page classification. *Arab. J. Sci. Eng.* **40**(12), 3555–3566 (2015)
3. Wang, J., Zhang, Y.: Opportunity model for e-commerce recommendation: right product; right time. In: SIGIR '13, pp. 303–312. ISBN: 978-1-4503-2034-4 (2013)
4. Vijayaragavan, P., Ponnusamy, R.: Socio economic psycho knowledge based intelligent agents for automated e-commerce negotiation. *HCI, LNCS* **6764**, 274–284 (2011)
5. Junior, J.M., dos Santos Moreira, E.: Using Support Vector Machines To Recognize Products In E-Commerce Pages. *ICMC*, 1–6 (2004)
6. Warmerdam, V.D., Szlyik, Z.: Confusion matrix based reweighting. *Contemporary Challenges and Solutions in Applied Artificial Intelligence, SCI* **489**, 143–148 (2013)
7. Joachims, T.: A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. *Universitat Dortmund* (1997)
8. Martins Junior, J.: Classificao de pginas na Internet, Dissertation (Mastering), Instituto de Ciencias Matemticas e de Computao Universidade de So Paulo, So Carlos, p. 81 (2003)
9. Joachims, T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features. *Universitat Dortmund* (1998)
10. Joachims, T.: Transductive inference for text classification using support vector machines. In: *Proceedings of International Conference on Machine Learning (ICML)* (1999)

ICI Reduction by Parallel Concatenated Encoder Using Wavelet Transforms



Noura Ali, M. I. Youssef and I. F. Tarrad

Abstract This paper proposed a new channel encoder technique by parallel concatenated encoders which use Hamming parallel with convolutional to reduce the “intercarrier interference” (ICI) by using the discrete wavelet DWT. The Hamming encoder is simple, and convolutional encoder gives low BER and robust against the burst error. Simulation the two encoder parallel via interleaver reduces the code rate and improves the performance. The simulation results measure the BER after applying the self-cancellation (SC) technique and appear reduction in BER in the results which enhance the system performance.

Keywords OFDM · PAPR · FFT · DWT · AWGN

1 Introduction

The difference between transmitter and recipient frequency in orthogonal OFDM system due to motion causes loss in orthogonality. The Doppler and orthogonality loss cause the “ICI.” The rate of frequency offset due to local frequency oscillator causes mediation. The undesired “ICI” degrades the system performance and causes a severe deterioration of quality of service in the systems. The techniques used to reduce the interference improve the system performance and mitigate its effect. The literature review as in [1] surveys the CS technique and its theory. In [2] the frequency offset using the self cancelation method. In [3] transform for the downlink multicarrier (DWT-MC-CDMA) is presented to mitigate these intrusions. Lastly,

N. Ali (✉)
Higher Technology Institute (HTI), 10th of Ramadan, Egypt
e-mail: nora.ali@hti.edu.eg

M. I. Youssef · I. F. Tarrad
AL-Azhar University, Cairo, Egypt
e-mail: mohiyosof@yahoo.com

I. F. Tarrad
e-mail: ibrahimtarrad@Azhar.edu.eg

in [4] operator perturbation technique is proposed for carrier mitigation in MIMO-OFDM.

1.1 Cancellation Method (SC)

The main concept of the self-cancellation scheme is to modulate the input data symbol onto several subcarriers. Allocate to these subcarriers predefined coefficients such that the resultant ICI signals cancel each other [5]. This method is more suitable for multipath fading channels because no channel estimation is required and the channel random [6].

1.2 Reducing ICI by DWTs

The conventional OFDM suffers from the interference deficiency in the received signal. System by DWTs combats better the narrow band mediation and considered as more robust to interference [7]. The simulation results use the DWT which replaces the sinusoidal function to wavelet function. The wavelet has some properties which enable it to reduce ICI in the OFDM system. Some of these properties are [8]:

- (1) The localization in time and frequency domain used in wavelet provides better orthogonality between subcarriers.
- (2) The basis wavelet functions are longer with an irregular shape which offers high degrees of suppression for sideband.
- (3) Wavelet does not use the cyclic prefix and save 20% from the system bandwidth.
- (4) The wavelet symbols have the property of double shift orthogonality which reduces ICI.
- (5) Wavelet analyzes the local properties of the input signal, such as edges or transients.
- (6) The application of self-cancellation in wavelet is the underwater acoustic communication [9].

2 Proposed Parallel Convolutional/Hamming Encoder

The proposed coding algorithm used is the Hamming encoder parallel with the convolutional encoder via the interleaver as presented in Fig. 1 [10].

In the proposed coding, the first encoder (Hamming) receives input bits directly [11], whereas the second encoder (convolutional) is fed with input bits through the proposed interleaver. The input bits ($x_k = x_1 \ x_2 \ x_3 \dots x_n$) is entered to the Hamming encoder and produced bit streams x_{kh} . The convolutional encoder receives the same

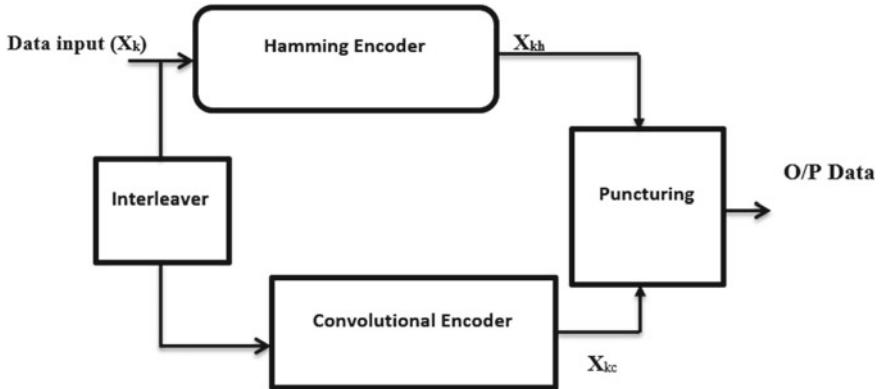


Fig. 1 Proposed coding algorithm

input data after passing through the proposed interleaver and produces bit streams x_{kc} . The output is produced when multiplexing the input data with the encoder's outputs x_{kh} and x_{kc} (parity bits). Puncturing is a way of removing or adding parity bits based on the channel properties and let the coded bits numbers fit the available bits in the physical channel (rate matching). This method also can be used to prioritize between systematic bits and parity bits.

The proposed channel coding algorithms uses the parallel concatenated levels. The convolutional (n_i, k_i) with a code rate of $r_i = k_i/n_i$, the Hamming encoder (n_o, k_o) with a code rate of r_o the final code rate $r = r_i * r_o$. Then, the parallel concatenated encoder/decoder levels increase the error correcting probability which is less than the two convolutional encoders and less complex and simulation time than the two convolutional encoders. The BER is calculated as in [12].

2.1 BER for OFDM Based on Wavelet Transform

The simulation results by MATLAB simulator use the BPSK modulation for 64 subcarriers to draw the BER versus SNR for conventional OFDM (OFDM-FFT) DW transform without Doppler ($D.$) shift as shown in Fig. 2.

Figure 2 shows the BER for conventional OFDM (OFDM-FFT) DW transpose. BER gives 0.006923 at 5 dB SNR without $D.$ shift. The wavelet does not need much SNR as FFT and also the BER is low. BER after applying SC scheme and parallel concatenated levels is clear in Fig. 3.

The lowest BER can be obtained required only 5 dB of SNR at a $D.$ shift of 0.3 as given in Fig. 3. BER versus SNR at zero $D.$ shifts is given in Fig. 4.

From Fig. 4, it can see that the system at zero $D.$ gives relatively lower values of BER after applying SC than before the SC scheme. BER before and after SC at $D. = 0.25$ are presented in Fig. 5.

Fig. 2 BER for OFDM-based DWTs

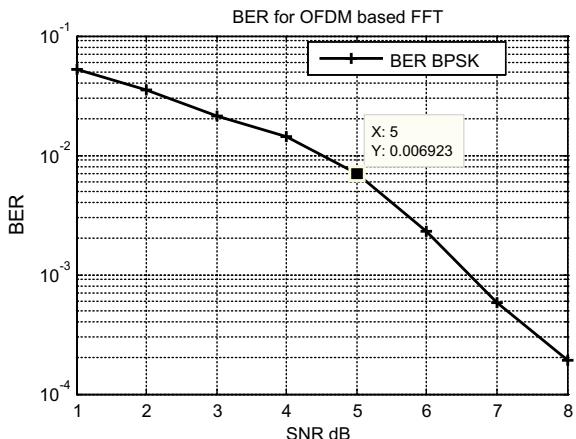


Fig. 3 BER for OFDM-DWTs after SC

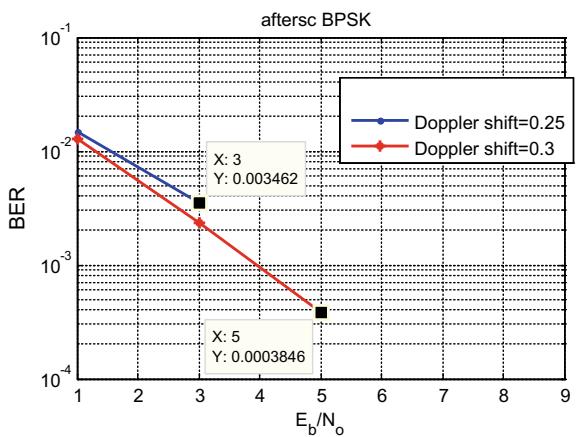


Fig. 4 BER after SC at zero D. shift

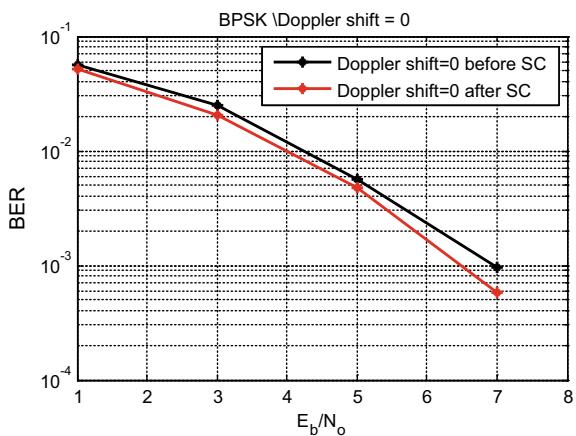


Figure 5 shows the BER when the $D.$ shift = 0.25 which gives 0.003846 at 5 dB SNR and gives 0.0001923 after applying the SC scheme to the system at the same SNR. The SC reduces a BER by a percentage enhancement of 96%. BER at a $D.$ shift 0.5 before and after SC scheme is clear in Fig. 6.

From Fig. 6, it deduces that BER at a $D.$ shift of 0.5 gives 0.005577 at 5 dB SNR while gives 0.0003846 at the same SNR. As the $D.$ shift increases, the BER increase. BER for 4-QAM after and before applying the SC scheme is shown in Fig. 7. Figure 7 shows that the BER is very high as compared to BPSK, and the SC scheme does not increase in performance.

Table 1 shows the BER for OFDM-based FFT and wavelet transpose by SC scheme. BER in OFDM-based DW transpose is less than of FFT transform in all values of SNR and $D.$ shift by a percentage enhancement nearly more than 97%. Comparison of the simulation results with the previous works is presented in Table 2.

Fig. 5 BER for OFDM after and before SC at $D.$ 0.25

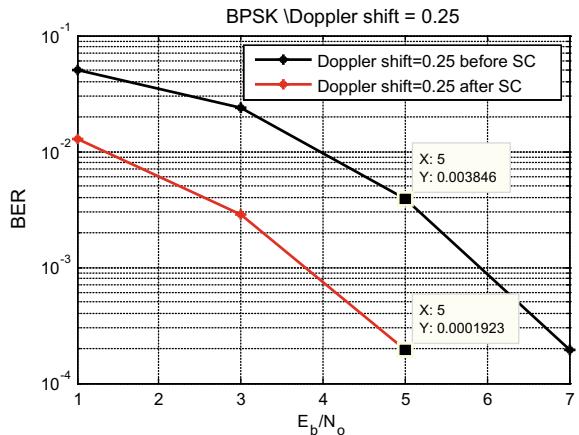


Fig. 6 BER before and after SC scheme at $D.$ shift 0.5

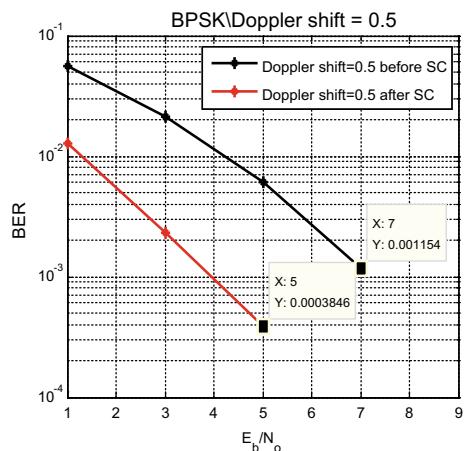
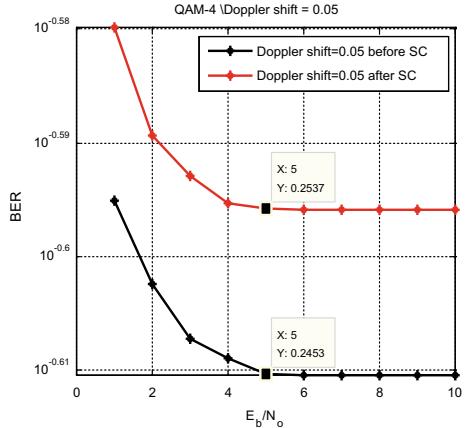


Fig. 7 BER at 4-QAM**Table 1** BER for OFDM-based FFT and wavelet transpose

Case studied	BER for OFDM-based FFT	BER for OFDM wavelet	% Enhancement
Standard BPSK	0.3733	0.006923	Nearly 98
Zero D. shift	0.3335	0.003269	97
D. = 0.25	0.3873	0.0001923	99
D. = 0.5	0.3785	0.0003846	99

Table 2 Comparison between the simulation results and the previous works

Case studied	BER for OFDM-based wavelet	Previous works	% Enhancement
Standard BPSK	0.006923	0.07 [1]2001	90.1
BPSK	0.003269	0.1 [2]2005	96.7
MQAM	0.001923	0.245 [5]2013	98
MQAM	0.001923	0.1 [5]2017	98

From Table 2, it can conclude that the BER is reduced for the proposed algorithm after applying the SC algorithm as compared with the mentioned previous works.

3 Conclusion

In the subsistence of small frequency offset and BPSK, self-cancellation gives the preferable results. Wavelet DWT transform gives less BER and hence good performance than FFT as described in Table 1. M-QAM is more sensitive to carry interfere. The wavelet-based system does not need a cyclic prefix to deal with delay spreads of the channel due to the overlapping nature of wavelet properties. Wavelet provides

side lobs of much lower magnitude than those of FFT and can better combat narrow-band mediation. The system is also inherently more robust with respect to interfere than traditional FFT OFDM and offers the simplicity of receiver design. The parallel concatenated levels lead to low BER and hence improve the performance. This paper applies the SC technique to mitigate the interfering effects using the DWTs by applying the parallel concatenated encoder for both Hamming and convolutional encoder. The final results give low BER for the proposed encoder and hence improve the system performance.

References

1. Gee, A.R.: A survey of ICI self-cancellation techniques for “OFDM systems”. *Int. J. Eng. Comput. Sci.* **4**(03) (2015)
2. Kumar, V., Singh, R.S.: Modified SSR ICI self-cancellation scheme for reducing bit error rate in “OFDM systems” (2014)
3. Ahmad, S., Hossam, H., Atef, G.: Interference cancellation for the downlink wavelet based multi-carrier CDMA system. *Int. J. Comput. Sci. Inf. Technol.* **8**(4) (2017)
4. Sherin, K., Abhiitha, E.: ICI mitigation in MIMO-OFDM by iterative equalization using OPT in time varying channels. In: 2017 International Conference on Intelligent Computing and Control (I2C2). IEEE (2017)
5. Z, Y, and. Hag. ‘IC interference self-cancellation scheme for OFDM mobile communication systems’ IEEE 49.7 (2001): 1185–1191
6. Chouhan, S., Sharma, D.: A survey of ICI reduction techniques in “OFDM System”. *Int. J. Comput. Trends Technol. (IJCTT)* **4**(8), 2677–2679 (2013)
7. Gauri, W., Ashok, K.: Performance analysis of FFT based “OFDM system” and DWT based “OFDM system” to reduce inter-carrier interference. In: 2015 International Conference on Computing Communication Control and Automation (ICCUBECA). IEEE (2015)
8. Jhin, A., Lavish, K.: Performance evaluation for wavelet based “OFDM system” affected by CFO over rayleigh channel. *Indian J. Sci. Technol.* **9**(5) (2016)
9. Kumar, Naresh, Sohi, B.S.: Wavelet based OFDM with ICI self-cancellation in underwater acoustic communications. *J. Comput. Sci. Inf. Secur.* **14**(4), 347 (2016)
10. Ali, N., Ibraheem, M., Fathi, I.: A proposed method to reduce the BER in wavelet based OFDM. *Int. J. Comput. Appl.* (0975 – 8887) **140**(2) (April 2016)
11. Ali, N., Ibraheem, M., Fathi, I.: A proposed method to reduce the BER in wavelet based OFDM. *Int. J. Comput. Appl.* **140**(2) (April 2016)
12. Ali, N., Ibraheem, M., Fathi, I.: Data compression on OFDM based on wavelet transform to enhance the LTE performance. IEEE (2018)

Sentence-Level Emotion Detection from Text Based on Semantic Rules



Dibyendu Seal, Uttam K. Roy and Rohini Basak

Abstract Detecting emotion from text has become an interesting topic in the field of natural language processing. Emotion detection aims to detect and recognize types of emotion from various sources such as text, facial expression and gestures, and speech. This paper proposes an efficient emotion detection technique by searching emotional words from a pre-defined emotional keyword database. The method analyzes the emotion words and phrasal verbs, also considers negation words and exhibits better performance than recent approaches.

Keywords Emotion detection · NLP · Natural language toolkit

1 Introduction

Emotion is essential to all aspects of human lives. It helps us in decision-making, affects our social relationships and shapes our daily behavior. Detecting emotion is helpful in human–computer interaction (HCI). With the wide use of social media like Twitter, Facebook, Blogs, product reviews and other public forum posts, there is a need to develop an algorithm to identify and analyze emotions available in text.

Emotion can be broadly categorized into six (6): happy, sad, anger, disgust, fear and surprise [1]. In this work, we have considered these six emotions to be detected from sentences which are supposed to consist of emotional words. Phrasal verbs and negative words have been analyzed to fine-tune the result. The model has been tested with the standard dataset and exhibits superior performance to recent works.

D. Seal (✉) · U. K. Roy · R. Basak
Jadavpur University, Kolkata, India
e-mail: dibyendu.seal@gmail.com

U. K. Roy
e-mail: royuttam@gmail.com

R. Basak
e-mail: visitrohinihere@gmail.com

Section 2 provides recent works in this area; Sect. 3 describes the emotion detection algorithm in detail, and in Sect. 4 we have listed the results and the issues we faced while implementation.

2 Related Work

Although there are various sources of emotion such as text, facial expression and gestures, and speech, detecting emotion from text is most challenging. Emotion detection from text relies on the emotional keyword analysis model.

In one of the primitive works [2] by Cecilia Ovesdotter et al., the authors tried to classify the emotional affinity of sentences in the children's fairy tales for future use of text-to-speech synthesis. In text-based emotion prediction (TEP) task, they tried to classify emotions from children's fairy tales using supervised machine learning and got an encouraging result in a first set of experiments.

As per Chomsky and Saussure [3], sentences can be broken down into two groups of words: function words and content words. Function words include nouns, adjectives, verbs and adverbs, while content words are prepositions, conjunctions and auxiliary verbs. Function words have the most influence on emotion of the text. These words supply basic emotion values or connection to the sentence where content words contain zero or minimum weightage.

Tao [4] analyzed the functional words and classified those into an emotional keyword, modifier words and metaphor words. Emotional keywords are those which can directly relate to the emotion, and modifier words increase or decrease or negate the emotion. Metaphor words are those which are not directly linked to an emotion. For example, 'kindness' always shows the gentle and positive emotion, 'joy.'

Ezhilarasi and Minu [5] tried to extract emotion using WordNet and its word construction. They collected the synsets (inner meaning) of the input from various domains to create an emotional ontology which is further used to classify complex sentences. However, their work toward creating the emotional ontology was limited to affected verb only.

Das and Bandyopadhyay [6] described the method of text emotion identification depending on emotion tagged word. They used WordNet for preprocessing and the conditional random field (CRF)-based word-level emotion classification. They proposed methods for handling negative words in sentence-level emotion detection.

3 Proposed Work

Possibility of extracting emotion or any other computational information from a sentence or text would not be very meaningful until some fundamental artificial intelligence is applied to text. The overall emotion extraction process broadly consists

of text preprocessing, keyword extraction from sentences and keyword analysis to find the emotional affinity of the keyword (Fig. 1).

3.1 Sentence Preprocessing

Text preprocessing performs the following tasks:

- Break the complex and compound sentence to a collection of simple sentences.
- Extract negative and high-intensity words from the sentence, e.g., ‘*Not*’ being considered as negative word whereas ‘*Very*’ being considered as intensity word.
- Parse the input text and extract necessary keywords using POS tagging.
- Extract phrasal verb from the sentence using trigrams and assign special tags to the phrasal verb in order to process them separately further.

Finally, the output of sentence processing module, i.e., extracted keyword, phrasal verbs, special emotion words, negation and intensity words, is passed to the keyword analysis module. On a high level, the module works as below (Fig. 2).

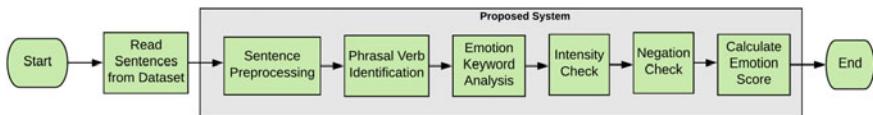


Fig. 1 Emotion detection process flow

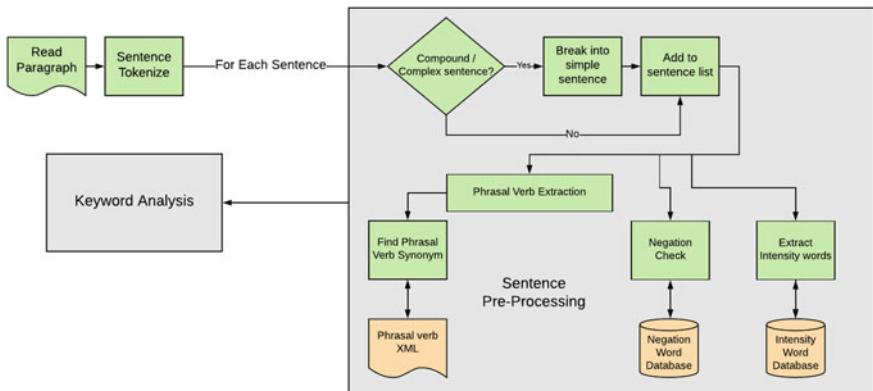


Fig. 2 Sentence preprocessing

3.2 Phrasal Verb Analysis

Phrasal verbs play a significant role in emotion detection. We analyzed the ISER emotion dataset and found around 50 phrasal verbs which can directly be assigned to emotion class. No other recent emotion detection work had taken phrasal verbs into consideration. <https://www.powerthesaurus.org> is an excellent resource of English words and their synonyms. We constructed an exhaustive list of phrasal verbs (around 1200 phrasal verbs) and created a database of synonyms of each phrasal verb. For example, ‘*Cheer up when the night comes, because mornings always give you another chance.*’—the phrasal verb ‘*Cheer up*’ gives a positive meaning of the sentence. The following phrasal verbs show different emotions:

Anger. Put back, lie around, come along, pass away, go around, go away, talk back, scare away, get back, give away, walk away, get along;

Fear. Pass away, move away, get away, go in, walk back, walk away, go back (Fig. 3).

The output XML contains the phrasal verb and their synonyms as given below (Fig. 4).

Extraction of PV from Text

There are two methods being used to extract the PV from the input text.

- Direct comparison of PV from the phrasal verb database is done to find out the words which are used ‘as-is’ in the text. If the match is found, we further look into the emotion database for matching synonyms in keyword analysis module.

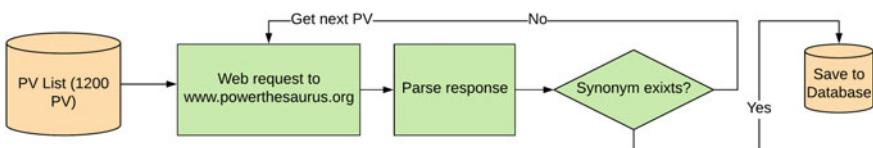


Fig. 3 Phrasal verb synonym extraction process

Fig. 4 Phrasal verb synonym

```

<PVList>
  <Verb>Account_for<syn>explain</syn></Verb>
  <Verb>Account_for<syn>justify</syn></Verb>
  <Verb>Account_for<syn> rationalize</syn></Verb>
  <Verb>Account_for<syn>elucidate</syn></Verb>
  <Verb>Account_for<syn>explicate</syn></Verb>
  <Verb>Account_for<syn>describe</syn></Verb>
  <Verb>Account_for<syn>attribute</syn></Verb>
  <Verb>Account_for<syn>clarify</syn></Verb>
  <Verb>Account_for<syn>number</syn></Verb>
  <Verb>Account_for<syn>cover</syn></Verb>
  <Verb>Account_for<syn>expound</syn></Verb>
  <Verb>Account_for<syn>furnish</syn></Verb>
  <Verb>Account_for<syn>impart</syn></Verb>
  <Verb>Account_for<syn> rationalise</syn></Verb>
  <Verb>Account_for<syn>solve</syn></Verb>
  <Verb>Account_for<syn>spread</syn></Verb>
  <Verb>Account_for<syn> vindicate</syn></Verb>
  <Verb>Account_for<syn> accredit</syn></Verb>
</PVList>
  
```

- Trigrams are useful when the phrasal verb is being used along with an additional word in between, e.g., your behavior has pissed me off. From this, pissed me off trigram matches with PV list and the synonyms are looked into emotion database for matching keyword.

The process replaces phrasal verbs and assigns special tag: *_EmotionPV_ <Seq No>*, for the subsequent process to understand the occurrence of phrasal verbs.

3.3 Emotional Keywords

First step of emotion detection is discovering keywords and phrases that associate with emotion. A list of words and corresponding emotion is called emotion lexicon. This list starts with root emotion words and is further extended using the WordNet synonyms. We added emotion synonyms to the list from WordNet synonyms and online dictionary (<https://www.powerthesaurus.org>).

The final list contained several hundred emotion words, for example.

Anger words. Furious, irritated, indignant, shout, mad, disturbed, annoyed, etc.;

Fear words. Anxiety, fright, worry, terror, dismay, horror, etc.

3.4 Keyword Analysis

Keyword analysis module takes the keyword from the sentence as input, looks for the keyword in the emotion lexicon and assigns a score. This module also gets synonyms for the phrasal verbs from phrasal verb synonym list. This module first does the stemming and POS tagging of the tokenized words to get root and restrict pronoun, preposition and interjections to get into the synonym lookup process as those words do not contribute to the emotion directly. This module considers noun, adjective, verb and adverb words for emotion detection keywords being searched into ‘emotional keyword’ lists and synonyms of phrasal verbs being searched from the phrasal verb xml, and the synonyms are further searched into ‘emotional keyword’ lists. The overall architecture of keyword analysis is given as (Fig. 5):

The emotion database contains a large number of words that affects the emotion directly. However, it is impossible to create a comprehensive list of such words. Not only that, many words indirectly represent to an emotion. For example, ‘Win’ does not define an emotion by its meaning, but any sentence containing ‘Win’ has a greater possibility of having joyous character rather than anything other. Our algorithm searches for the key-word in the emotion database and if keyword being matched, then we assign a point to the corresponding emotional score. If the match is not found, then we first extract the synonyms of the word from WordNet lexical database [7] and further search synonyms in the emotion database. We repeated this process for three levels to get significant coverage of keywords. However in each level, we reduced

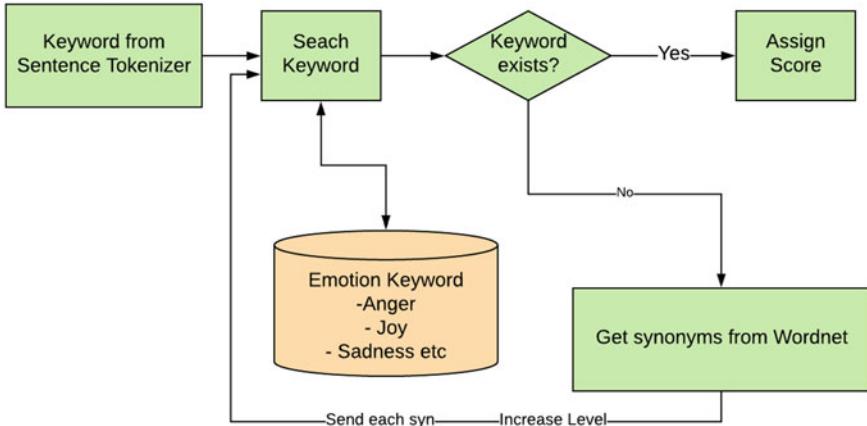


Fig. 5 Keyword analysis

the weight by 33% as meaning gets changed significantly in each level, and beyond third level, the process yields erroneous results.

```

Iteration = 0
Weight = 1
function CheckEmotion:
for keyword in keyWordList:
  If Iteration < 3
    Iteration ++
  Check keyword exists in Emotion Database
  Yes:
    Emotion Score += Weight (Full weightage given)
  NO:
    Get synonyms List
Weight = Weight*0.66 (33% reduction)
GOTO function
  
```

The emotion score is calculated as below:

$$\text{Emotion}_{\text{angry}} = \sum \text{Weight of Angry for All the Keywords}$$

$$\text{Emotion}_{\text{sad}} = \sum \text{Weight of Sadness for All the Keywords}$$

$$\text{Emotion}_{\text{joy}} = \sum \text{Weight of Joy for All the Keywords}$$

$$\text{CalculatedEmotion} = \text{Max}(\text{Emotion}_{\text{angry}}, \text{Emotion}_{\text{sad}}, \text{Emotion}_{\text{joy}})$$

Table 1 Effect of negative and intensity words on emotions

Emotion	Negative words	Intensity words	Both
Angry	Neutral	Disgust	Neutral
Sadness	Neutral	Sadness	Neutral
Fear	Fear	Fear	Fear
Joy	Anger	Joy	Anger
Surprise	Sadness	Fear	Neutral

3.5 Intensity Negation and Check

Emotion detection from the text algorithm considers negative and intensity words in sentence and changes the assigned score derived from the keyword analysis. Negative word changes the emotion significantly, whereas intensity words intensify the meaning. A list of negative and intensity keywords is created, and we find the occurrence of the words in the list.

Negative words. No, not, negative, neither, nor, least, none, nothing;

Intensity words. Very, big, deep, high (Table 1).

4 Experimental Results

We used emotion dataset provided by ‘International Survey on Emotion Detection Antecedents and Reactions’ (ISEAR) containing around 7500 sentences affecting anger, joy, sadness, fear, disgust and guilt. We created individual emotion dataset containing emotional keywords and observed an overall accuracy of 65%. The current model computes ‘neutral’ emotion to provide more realistic output. ‘Neutral’ emotion was not classified in ISEAR dataset and that also reduced the accuracy. To justify the consistency and correctness of the algorithm, we validated our result w.r.t to the result obtained by Das et al. [6].

a = no. of sentences correctly assigned to the emotion;

b = no. of sentences incorrectly assigned to the emotion;

c = total number of sentences;

d = no. of sentences with ‘neutral’ emotion (Table 2).

4.1 Analysis and Discussion

The stemming algorithms are not robust. Porter stemmer failed to stem certain words. For example, ‘startle’ (which means surprise/fear) becomes start and ‘Sorry’ becomes ‘Sorri’. WordNet synsets are not proper in many times. For example, synonym of

Table 2 Test result comparison

Metrics	Expression	Current experiment result (%)	Conditional random field framework [6] (%)
Accuracy	a/c	65	60.47
Precession (P)	$a/(a + b)$	70.11	67.95
Recall (R)	$(a + d)/(a + b + d)$	62.67	65.11
F1 score	$2*(\text{Recall} * \text{Precision})/(\text{Recall} + \text{Precision})$	66.18	Not found

‘Last’ is death which belongs to fear emotion. Few sentences are not grammatically correct; rather, it was more aligned toward the style of speaking of an individual. Current algorithm searches each of the synonyms in emotion database. However, there should be a suitable mechanism to select the orders of those words. Idioms that describe an emotion could not be constructed in this experiment.

5 Conclusion

This paper describes the emotion detection process by searching keywords from emotion database. The proposed system improves the performance over existing methods, though not all the limitations of the existing systems are resolved. The emotion database still does not contain an exhaustive list of emotional keywords, and we are still analyzing more resources to improve accuracy.

References

- Ekman, P.: Emotions Revealed: Understanding Faces and Feelings, 1st edn. Weidenfeld & Nicolson (8 May 2003). ISBN-13: 978-0297607571
- Alm, C.O., Roth, D., Sprout, R.: Emotions from text: machine learning for text-based emotion prediction. In: HLT '05 Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 579–586
- Saussure and Chomsky: The Distinctness of their ‘Identical’ views, Magyar Nyelv (Hungarian Language), pp. 420–430 (2006)
- Tao, J.: Context Based Emotion Detection from Text Input. Interspeech 2004, National Pattern Recognition Laboratory, Chinese Academy of Sciences, pp. 1337–1340
- Ezhilarasi, R., Minu, R.I.: Automatic emotion recognition and classification. In: international conference on modelling, optimization and computing (ICMOC–2012), pp. 21–26
- Das, D., Bandyopadhyay, S.: Sentence level emotion tagging: 2009. In: 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 145–162. <https://doi.org/10.1109/aci.2009.5349598>
- WordNet. [Online]. Available: <http://wordnet.princeton.edu/>

A Real-Time Automatic Human Facial Expression Recognition System Using Deep Neural Networks



Surekha Samsani and Vineel Abhinav Gottala

Abstract Nowadays, there is a great need for human–machine interaction (HMI) systems with communication channel between humans and machines to share a variety of tasks. To communicate emotions and expressions for humans, facial expression is the easiest way and also being capable of recognizing facial expression could lead to more responsive and intelligent systems that might improve the user experience. This paper presents a real-time automatic human facial expression recognition (RTA-HFER) system, which recognizes the expression of humans in a streaming video using the most popular deep learning convolutional neural networks (CNN). The RTA-HFER system is trained on 32,298 training records taken from Kaggle challenges repository, and to accelerate its performance and scalability, the graphical processing unit is used. The results revealed that the classification accuracy and training time of RTA-HFER system are better when compared against inception network, VGG-16 and self-designed CNNs.

Keywords Deep learning · Human–machine interaction · Facial expression recognition · Convolutional neural networks · Kaggle challenges repository

1 Introduction

In the present generation, people are depending on the technologies to fulfill their tasks more efficiently and also in a better way specifically to get round the clock services. During recent years, a lot of researches are going on to develop intelligent systems that could be more responsive and less interactive; i.e., there is a great demand for human–machine interactive systems [1]. The main idea of developing intelligent

S. Samsani (✉)

JNTUK-University College of Engineering Kakinada, Kakinada, India

e-mail: surekha.samsani@gmail.com

V. A. Gottala

JNTUK-University College of Engineering Vizianagaram, Vizianagaram, India

e-mail: vineelabhinav21@gmail.com

interactive systems is to perceive, act, and make necessary decision according to the current situation and take the best action with less human interference. Some of the applications of such intelligent interactive systems in real world include intelligent analysis, military, hospitality industry, purchasing technologies, smart classroom, tele-teaching, smart driving, lie detector, driver alertness, video games, etc. A smart environment in effective learning, hospitality industry, and purchasing technologies is much more helpful to provide efficient services by enabling intelligent communication between humans and computer technologies; i.e., HMI is more relevant and helpful for service industries. One of the major technologies in HMI is the facial expression analysis [2] and recognition of emotion, because facial expression analysis can be used in wide variety of applications like tele-teaching, hospitality, purchasing, smart driving, neuroscience, linguistics, psychology, pain monitoring, and other related disciplines. Face emotions play a vital role in understanding the behavioral aspects of humans in a social intelligent communication, and analyzing the facial expression of humans will be more helpful for the intelligent systems to provide better service and also to make the best decision at a situation. Supervised machine learning techniques [3] provide the computers the ability to learn automatically from training data and apply the gained knowledge to make better decisions without much human intervention. Hence, the most emerging machine learning techniques can be applied to develop an effective facial expression classification system, and deep neural networks [4] are the most advanced machine learning-based techniques being applied successfully in the area of computer vision.

Deep neural networks (DNNs) mimic the working of human brain, and generally, it has one input layer followed by at least one hidden layer and an output layer. Multiple hidden layers in DNNs help in extracting low-level features (like edge detection), which will be helpful to detect higher-level features. In facial expression recognition task, it mainly deals with image data which will have several low-level features (detection of edges), medium-level features (detection of eye region), and high-level features (detection of whole face which is helpful in finding expression). This kind of different level of extraction needs more number of hidden layers as in deep neural networks. Compared to other techniques like artificial neural networks, support vector machines, and principal component analysis, deep neural networks do not require feature extraction because through deep networks it will automatically extract features which are helpful in recognizing a particular expression. For better performance and accuracy, DNN requires high computational power and more number of training examples compared to other techniques. This paper proposes a real-time automatic human facial expression recognition system (RTA-HFER) based on deep learning convolutional neural networks (CNN) [5–8] to analyze the seven universal human expressions like happy, neutral, sad, fear, disgust, surprise, and anger. The precision and accuracy of the RTA-HFER system is tested on the benchmark human face image dataset taken from Kaggle competition challenges repository [9].

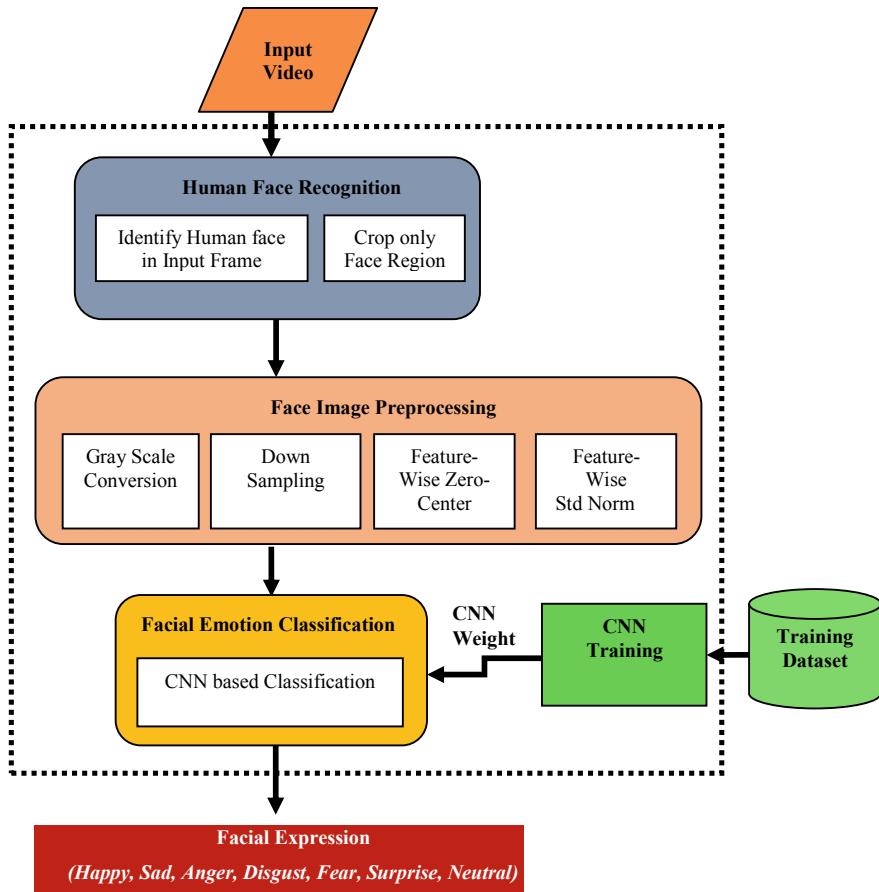


Fig. 1 RTA-HFER system

2 Methodology

The proposed real-time automatic human facial expression recognition system (RTA-HFER) is about to detect and analyze the human expressions like happy, sad, disgust, anger, fear, surprise, and neutral in a given input video.

The RTA-HFER system operates by detecting the human face in the given video and then applies some pre-processing techniques on the detected human face image in order to recognize the human emotion accurately and then classifies the expression of the human face. The methodology of the RTA-HFER system is given in Fig. 1. So, the overall methodology of the RTA-HFER system can be divided into three phases, namely

Phase-1: Detect human face in the input video

Phase-2: Pre-processing of the facial image

Phase-3: CNN-based facial expression classification.

Phase-1: Recognize human face in the input video

The overall goal of RTA-HFER system is to recognize the emotion of human face in the input video stream. Hence, the very first step in the RTA-HFER system is to detect the position of human face in the input video frame. So, OpenCV Haar cascade-classifier [10] is used to detect human face in the input frame and segment the face image so that it will contain only the face part of human being and remove rest of the background information which is not essential to the expression classification. The cropping region also removes the other facial parts that do not contribute to the expression.

Phase-2: Pre-processing of Human face image

The following is the sequence of techniques to be carried out on the face image.

1. Grayscale conversion
2. Down sampling
3. Feature-wise zero center
4. Feature-wise standard normalization.

As the color is not an essential attribute to identify the emotion of a human face, collapse the RGB channels into Grayscale channels. Converting the cropped human face image to a grayscale channel not only reduces the computational power but also helps in improving the performance of the classifier. The size of the image also affects the performance of the RTA-HFES system. So, the next pre-processing step is to downsample the face image to 48*48 size. The final step in image pre-processing [11] is to normalize the data to ensure that each input parameter has a similar data distribution. So, feature-wise zero center and then feature-wise standard normalization are done to normalize the pixel data such that all pixel data will have equal distribution.

Phase-3: CNN-based facial expression classification

After pre-processing the human face image, it is to be submitted to the CNN-based classifier for recognizing the emotion of the face.

CNN-based facial expression classification is an important phase in the RTA-HFES system. The overall efficiency of the RTA-HFES system depends on the classifier used to detect the expression. Before being applied the image for classification of expression, convolutional neural networks should be trained to obtain the CNN weights. Hence, this phase can be viewed as three sub-phases. They are:

1. Train the convolutional neural network(CNN) with training dataset
 2. Store the obtained weights and biases (parameters) in file system
 3. Feed the pre-processed input image to CNN for classification.
- 1. Train the convolutional neural network (CNN) with training dataset.** The CNN architecture of the proposed RTA-HFER system is similar to ResNets [12], and the basic building block of ResNet is residual block and is shown in Fig. 2.

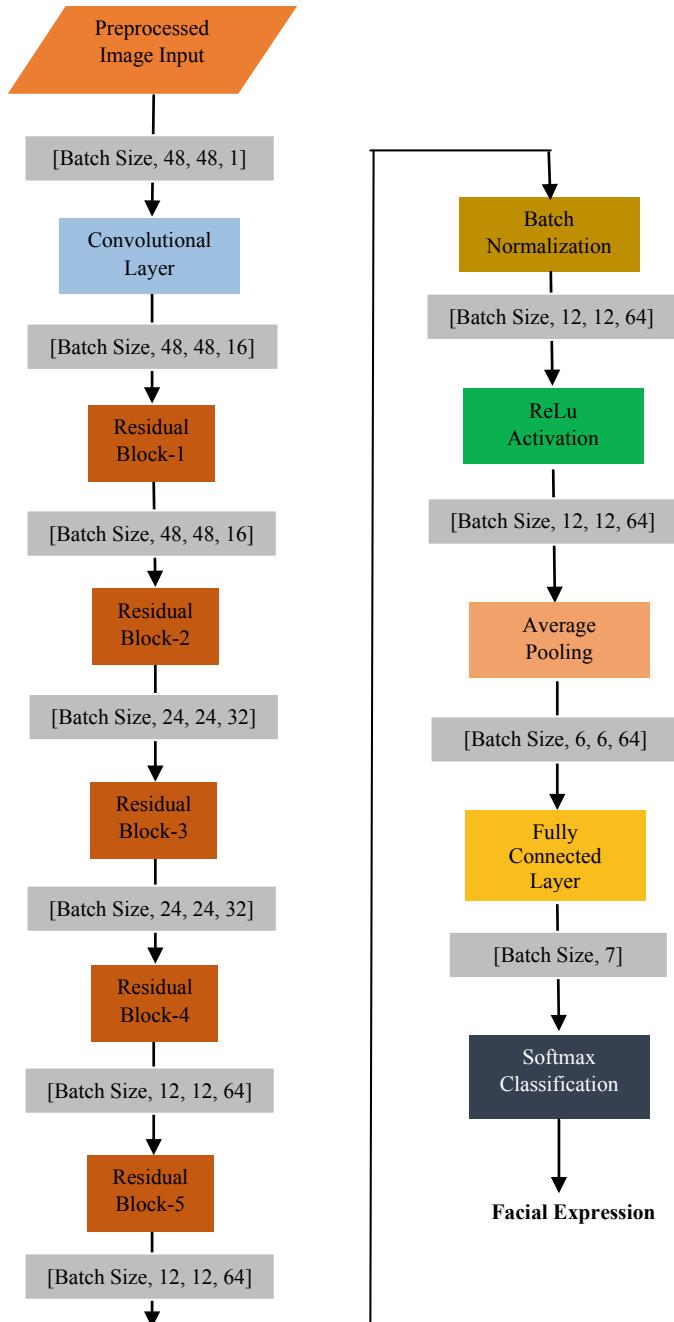


Fig. 2 CNN architecture of the proposed RTA-HFER system

Convolution Layer (CONV Layer). The convolution layer [6, 7, 13] applies a set of kernels (also called as learnable filters or neurons) over the input images. The depth of these kernels should be same as that of the depth of input image. Every kernel is an array of random values, which slides over the input image at every position to compute the dot product between the parameters in kernel and the input image and produces a feature map or activation map.

Pooling Layer. The pooling layer [6, 7, 13] operates on each feature map produced by the convolution layer, and its function is to minimize the spatial dimensionality. This downsampling reduces the computation and also helps to overcome overfitting problem. The two most common pooling layer operations are:

- MAX pooling—maximum of values will be taken from window
- AVERAGE pooling—average of values will be taken from window.

ReLU Activation Function. Rectified linear unit (ReLU) [6, 7, 13] is an activation function which is used to remove the negative values in input activation. The output of a rectified linear unit can be computed as,

$$\begin{aligned} F(x) &= 0, \quad \text{if } x < 0 \\ &= x, \quad \text{otherwise} \end{aligned} \tag{1}$$

Linear Activation Function: It is also an activation function where the activation is proportional to input.

Residual Block

Residual blocks [12] are the basic building blocks used in residual network architecture.

Figure 3 represents a basic residual block where x is input, $F(x)$ is the residue from in between layers, and $H(x)$ is the output. This block is a general basic structure where it has only two layers named weights. The residual block structure used in the proposed RTA-HFER is similar to Fig. 5 with combinations as follows,

- Conv Layer + Linear Activation Function when downsample is false
- Conv Layer + Linear Activation Function + Max Pooling when downsample is true.

The Hyper-parameters filter size, number of filters, and strides are taken randomly for each layer in the residual block. With the above combinations, different variants of residual blocks are possible. Stacking up of these residual blocks, any number of

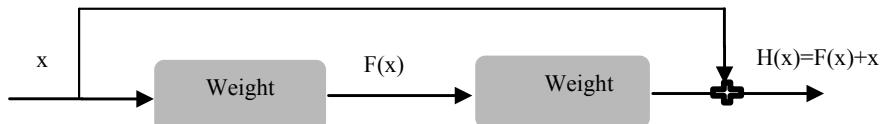


Fig. 3 Residual block

Table 1 Configuration of residual blocks used in RTA-HFER system

Residual block number	No. of filters	Filter size	No. of blocks	Downsample
Block-1	16	3×3	5	False
Block-2	32	3×3	1	True
Block-3	32	3×3	4	False
Block-4	64	3×3	1	True
Block-5	64	3×3	4	False

times forms the residual neural networks. The configuration of the residual blocks used in RTA-HFER system is given in Table 1.

Fully Connected Layer: Like in regular neural networks, the FC layer [8] has full connections to all activations in the previous layer. All the feature maps from the previous layer are combined in this layer to form a filter for each classification and these filters will be applied on the test image. Softmax regression [14] is used for classification with seven neurons representing the seven emotions to predict the class of the given input.

2. **Store the obtained weights and biases (Parameters) in file system.** After training the CNN model with train database, the obtained parameters like weights and biases are to be preserved to classify the images into respective classes. Since training time is very long, we need to save these weights so that we can use later during testing phase.

During testing phase, we load the parameters from hard disk and use it on the test database or test image which classifies the images into correct classification. This sub-phase saves the following four files into file system.

- Checkpoint file
- Metafile
- Data file
- Index file.

Checkpoint file: Checkpoint file keeps a record of latest files created, which is the last saved model. It is like restoration point.

Metafile: Metafile saves the complete tflearn graph, i.e., all variables, operations, collections, etc. This file has meta-extension.

Data file: Data file contains all the values of weights, biases, gradients, and all other training variables saved.

Index file: Index file is string-string immutable table. Each key is name of tensor, and its value is serialized BundleEntryProto. Each BundleEntryProto describes metadata of tensor like which of data file contains content of tensor, the offset into that file, etc.

3. **Feed the pre-processed input image to CNN for classification.** In this phase, the final pre-processed human face image is submitted to the CNN-based classifi-

cation, where it applies the pre-defined parameters and gets the final classification of emotion.

3 Experiment and Results

The RTA-HFER system is implemented in Python 3.5.0 version (dependencies are TensorFlow, TFlearn, Numpy, Pandas, Matplotlib, Sklearn, OpenCV) on a computer with Intel Core i7-7700HQ processor 2.6 GHz, 8 GB RAM with NVIDIA GeForce GTX 950 M with 4 GB dedicated memory.

3.1 *DataSet*

For experimental analysis, the FER dataset has been taken from the Kaggle competitions repository [9], and the data consists of 48×48 pixel grayscale images of faces. The labels and their respective classes are given below,

Angry – 0,
Disgust – 1,
Fear – 2,
Happy – 3,
Sad – 4,
Surprise – 5, and
Neutral – 6.

The total samples taken from the Kaggle competitions repository have been divided into training data and test data. The training data set consists of 28,709 examples and 3589 examples for public test.

3.2 *Experiment*

The training of the network is achieved by momentum optimizer with a learning rate starts from 0.1 and momentum value 0.9 iterated through 100 epochs with a batch size of 100.

3.3 Results and Evaluation

The performance of the RTA-HFER system was benchmarked by conducting a series of experiments on the FER dataset. The obtained confusion matrix is shown in Fig. 4.

The performance metric's like precision, recall, and accuracies observed from confusion matrix and the obtained results are shown in Figs. 5, 6, and 7.

The performance of the proposed system is also compared with other methods [15], and the comparative analysis is shown in Fig. 7.

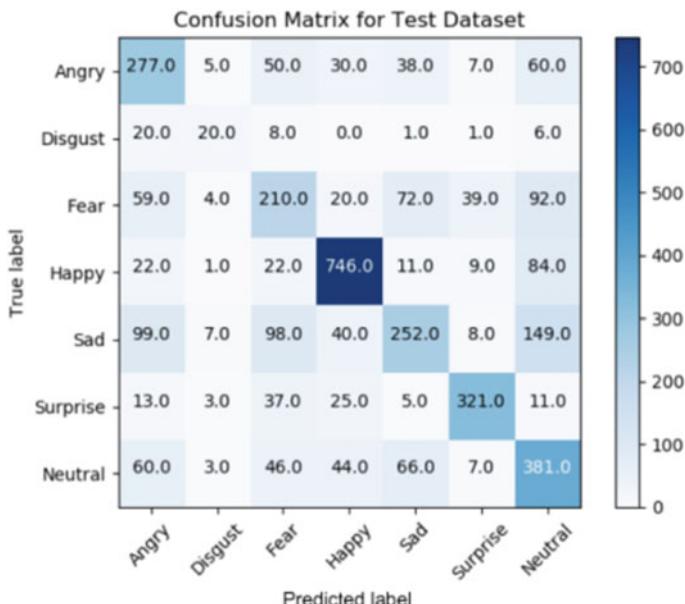


Fig. 4 Confusion matrix for 3589 test samples

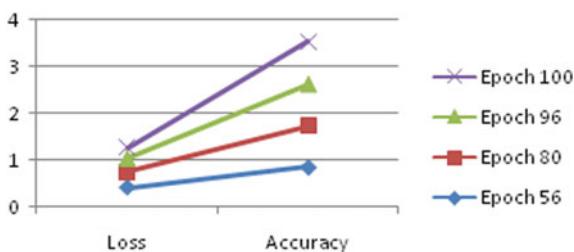


Fig. 5 Loss and accuracies obtained at different Epoch's on FER dataset

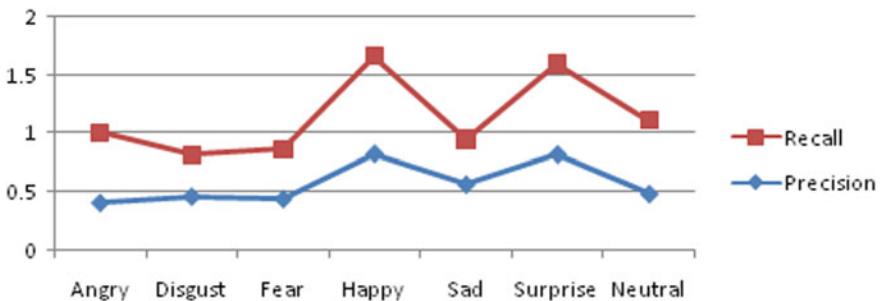


Fig. 6 Basic comparative measures of universal facial expressions on FER dataset

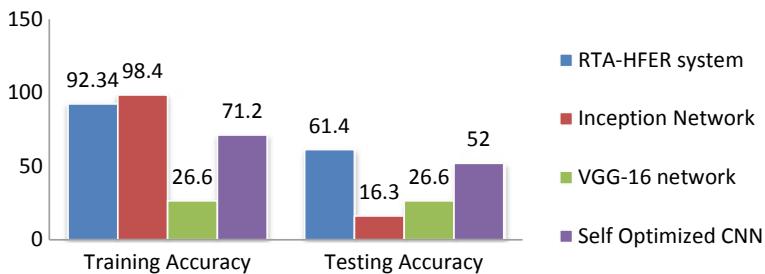


Fig. 7 Accuracies of different models on FER dataset

4 Conclusion

In this paper, an efficient and faster method to recognize the facial expressions is proposed based on residual network architecture. The RTA-HFER system uses residual blocks for identifying the most prominent features that constitute to a particular expression. Training and feature extraction using a residual network in deep learning framework tflearn is done on CUDA-enabled GPU, which reduced the training time significantly. Experiments were conducted on publically available dataset from Kaggle challenges which got a testing accuracy of 61.4% with training accuracy of 92.34%, which is comparably better than self-designed CNN, inception networks, and VGG-16 networks. And due to the lack of availability of train data per classification (i.e., for each emotion), residual networks could not obtain test accuracy not more than 61.4%. Moreover, every emotion of humans (represented via their faces) has overlapping features with one another, and hence, the probabilities of portraying one emotion overlap with other. So, the improved test accuracy of the proposed system when compared against other contemporary methods may help in recognizing the seven universal human facial expressions much more accurately.

References

1. Dix, A.: Human-computer interaction. In: LIU, L., ÖZSU, M.T. (eds.) Encyclopedia of Database Systems. Springer (2009)
2. Christine, L.L., Diane, J.S.: Automatic facial expression interpretation: where human-computer interaction, artificial intelligence and cognitive science intersect. *Pragmat. Cogn.* **8**, 185–235 (2000)
3. Mitchell, T.M.: Machine Learning. McGraw-Hill, Singapore (1997)
4. Fasel, B.: Multiscale facial expression recognition using convolution neural networks. In: Proceedings of the Third Indian Conference on the Computer Vision, Graphics and Image Processing, Ahmedabad, pp. 1–9 (2002)
5. Kotsia, I., Pitas, I.: Real time facial expression recognition from image sequences using support vector machines. In: IEEE Conference On Image Processing, vol. 3 (2005)
6. Karpathy, A.: Convolutional neural networks for visual recognition. Available online www.cs231n.stanford.edu
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
8. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
9. Kaggle Challenges Dataset.: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data>
10. Chandrappa, D.N., Akshay, G., Ravishankar, M.: Face detection using a boosted cascade of features using Open CV. In: Venugopal, K.R., Patnaik, L.M. (eds.) Wireless Networks and Computational Intelligence. Communications in Computer and Information Science, vol. 292, pp. 399–404 (2012)
11. Yamada, M., Jitkrittum, W., Sigal, L., Xing, E.P., Sugiyama, M.: High-dimensional feature selection by feature-wise kernelized lasso. *Neural Comput.* **26**, 185–207 (2014)
12. Residual Network Architecture.: http://github.com/tflearn/tflearn/blob/master/examples/images/residual_network_cifar10.py
13. Rawat, W., Wang, Z.: Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput.* **29**, 2352–2449 (2017)
14. Ren, Y., Zhao, P., Sheng, Y., Yao, D., Xu, Z.: Robust softmax regression for multi-class classification with self-paced learning. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17) pp. 2641–2647 (2017)
15. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-ResNet and the impact of residual connections on learning. *InAAAI* **4**, 4278–4284 (2017)

Ballast Water Quality Compliance Monitoring Using IoT



K. Komathy

Abstract Pumping sea water from the ballast tank is inevitable so as to stabilize the ship at every time the freight is uploaded or unloaded. However, this water holds a variety of organisms that embrace living species, microbes and viruses that belong to the coastal and marine biome. These organisms become alien when the ballast is discharged into the sea. These Martian species can bring hazards to ecosystem and biodiversity apart from environmental, social and financial impacts on the water region where the ballast water is discharged. Therefore, it is important that each ship shall invest in a ballast water treatment system as per the guidelines laid out by the statutory agencies. A worldwide regulatory is required to inculcate a uniform compliance on the treatment of ballast water. This paper, as an attempt towards this direction, explores the regulations for ballast water treatment system and studies the methodologies used for the treatment and proposes an automatic and intelligent system, which continuously monitors the compliance of the treated water using internet-of-things (IoT).

Keywords Ballast water treatment · Invasive alien species · Filtering · Ultraviolet treatment · Internet of things · Big data analysis

1 Introduction

Shipping sector is the backbone for the international trade merchandising over 90% of goods around the world. As the cargo ships depend on ballast tanks for safety and fuel efficiency, it is unavoidable that these ballast water tanks acting as an outlet for the transfer of species that are not native to the receiving water sources. These alien species pose a threat to aquatic culture, agriculture, fisheries and international trade. Various research and development works are funded to control and repair the damage these species cause. Therefore, the objective of the paper is to investigate

K. Komathy (✉)

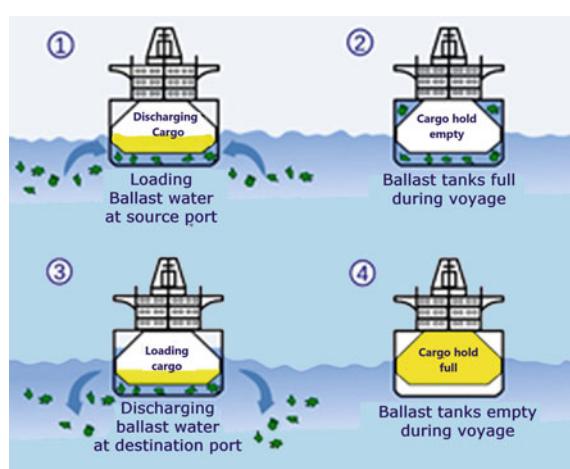
Academy of Maritime Education and Training, Chennai, India
e-mail: ithod@ametuniv.ac.in

the importance of ballast water treatment, the necessity to monitor the water quality and to propose an architecture for inspecting the compliance of ballast water that discharged out in the ocean in real time rather to wait for the data be submitted by the ships. The focus on this topic is ensued as to aid the regulatory body to monitor 3–5 billion tons of ballast water transferred by cargo ships per annum worldwide.

International Maritime Organization (IMO) introduced a convention for ballast water management (BWM) in 2004 [1], which devised more controlling measures against endangering the environment by the ballast water discharged into the sea. Towards the treatment of ballast water, this convention has brought out two significant standards namely “ballast water exchange (BWE) and ballast water performance” and these regulations are put into practice from September 2017. To keep the ballast water and sediments to the amenable level, each merchant ship is necessarily to monitor its approved plan, which carries out the treatment processes accounted in the ballast water management.

Figure 1 illustrates the various phases involved as a ballast water cycle. BWM plan [1] is expected to record the observations during treatment process, water quality measures, sampling techniques, investigations carried out, inferences arrived and unfavorable cases and challenges handled are maintained in the record. Location of the ship and ocean area range covered for ballast water exchange are also logged. The plan too includes the water regions where ships should not exchange water. The ships are expected to submit the strategies, policies and schedules used for BWM to IMO. Ships following the BWM standard have to restrict the number of various organisms acceptable in cubic meter of water when the ballast water is drained into the sea or ocean. In summary, the data in the record book shall reveal the salient information of BWM such as the schedules planned for ballast water treatment (BWT), GPS longitude and latitude of the ship, capacity of the ballast water tank verses cargo capacity, temperature, salinity and volume of ballast water uploaded and unloaded.

Fig. 1 Loading and unloading of water to ballast the ship. *Source* Hielscher—Ultrasound Technology



2 State-of-Art on BWM

Many research groups and technology developers have focused their works primarily on BWT and BWE. US regulation [2] emphasizes the monitoring policies towards reducing the detrimental effects of ballast water transferred onto the sea. Banerji et al. [3] developed a comprehensive approach for evaluating the BWM system against endangering the lives of mankind. The model assesses the type of BWM practiced, the usability and permissibility of water treated and the noxiousness of dissolved compounds of ballast water before it is discharged into the sea. California State Lands Commission [4] directed that the recently constructed vessels shall submit a detailed evaluation strategies for BWT system erected in the ship having a tank capacity of 5000 metric tons minimum. An effective treatment system also depends on the vessel type, shipping routes and port geography.

Emerging risks from BWT are investigated by Barbara et al. [5] and their findings are derived based on the mathematical design and physical, chemical properties and toxicity data. David [6] presented a new model for evaluating ballasting and de-ballasting processes and described the regulatory procedures for BWM. Elizabeta Briski et al. [7] have outlined the policy to manage ballast water combining both BWE and BWT systems in ships. This study examined the impact of the combined model against the individual model BWT or BWE alone on living organisms.

The BWM convention [1] guided the ships on the technical standards to govern and regulate ballast water and sediments. It instructs that all ships carry a book to record BWM procedures as per the standard. As per IMO 2008 [8], when a designated surveyor finds out that the BWM does not follow the steps as specified in regulation E-2 or E-3 of IMO 2004 [1] or if the ship poses a threat to the environment, human health, business, then necessary remedial measures are warranted immediately.

The goal of the guidelines given in another report by IMO 2008 [9] is to estimate the concentration of compounds and their impact in BWM system. The strategies provided in the IMO 2009 [10] ensures that the risks are to be stated clearly by the respective supplier who prepares active substances and as well as by the ship owner on the BWT system employed in the ship. Document [11] describes “the methodology for information gathering and the conduct of work of the Group of Experts on the Scientific Aspects of Marine Environmental Protection—Ballast Water Working Group (GESAMP-BWWG) when undertaking technical evaluations in accordance with the procedure for the approval of ballast water management systems that make use of active substances (G9), as revised and adopted by resolution MEPC.169(57)”.

The Indian register of shipping [12] has channeled the current state of BWT regulations and technologies to guide the ship owners, operators, and ship builders. The Danish environmental protection agency has produced an order [13] that covers ships flying the Danish flag in Danish territorial waters or in the exclusive economic zones. It stated that the amount of the water exchange to be made shall be equivalent to thrice the volume of ballast water tank taken from the port calls. Lloyd's register [14] covers the strategies used in the current BWT technologies. Lothar Schillak et al. [15] presented a model for (i) an on-board ballast water sampling system and (ii)

analytical on-board methods, which generate reliable data from the BWT system. Murphy et al. [16] inspected the active compounds in the ballast tanks of ships operating in the North Pacific and Atlantic oceans and found that the distance from land from where ballast water was drawn could be mapped to the concentration of elements.

Zofia Jóźwiak et al. [17] have shown that aquatic organisms from ballast water unloaded into Szczecin harbor could bring threats to the existing condition such as (i) creating food competitors for the local aquatic species, (ii) endangering human health by spreading unknown diseases to the local environment, and (iii) reducing fish resources. Study of an automated remote monitoring of ballast water given by Goran Bakalar et al. [18] revealed that remotely operated system from land improves traffic environment as the system verifies the contents of ships' ballast water before a ship enters a port. In another report, the authors [19] have necessitated a continuous supervision of the BWT system. The samples of the treated ballast water were analyzed using their proposed supervisory model wherein the warning messages are sent if the samples do not satisfy the permitted range or if BWT fails.

3 On-Board Ship Ballast Water Treatment System

Aleksandar et al. [20] concluded that the various parameters of BWT concerning (i) the category and model of the ship (ii) ballast water tank design and its capacity and (iii) the seas where the cruise travels and as well as harbors where ship exchanges the water. Some of the countries such as USA, Canada, Australia and New Zealand adhere their own BWM plans, which are more stringent than that of BWM convention. The traditional methods involved in ballast water treatment are basically categorized into (a) particle separation using mechanical methods (b) chemical treatment by active substances and (c) combination of above treatments.

Figure 2 displays the design of an on-board BWT chamber. Sea water pumped in the ballast water tank has to undergo a series of filtering for the compliance of water quality before it is drained into the sea. Online ultraviolet transmission (UVT) is currently adopted technique that compares how much light passes through a water sample versus a pure water sample at 254 nm expressed at a 1 cm path length. This measurement is used for UV control to allow energy savings by adjusting the output to match water quality and flow in real-time. The sensors measure Spectral Absorption Coefficient (SAC) and UV transmission (UVT) at 254 nm for determination of organic loads in water. With the sensor UV SAC, real-time turbidity can be detected directly without any chemicals. It comprises a controller module that includes an USB interface and an internal data logger up to 20 sensors for measuring parameters such as O₂, NH₄, NO₃, COD, PO₄ and sludge level. A communication module comprising an Ethernet-interface or a WiFi connectivity to processing module where the water quality is analyzed and reported.

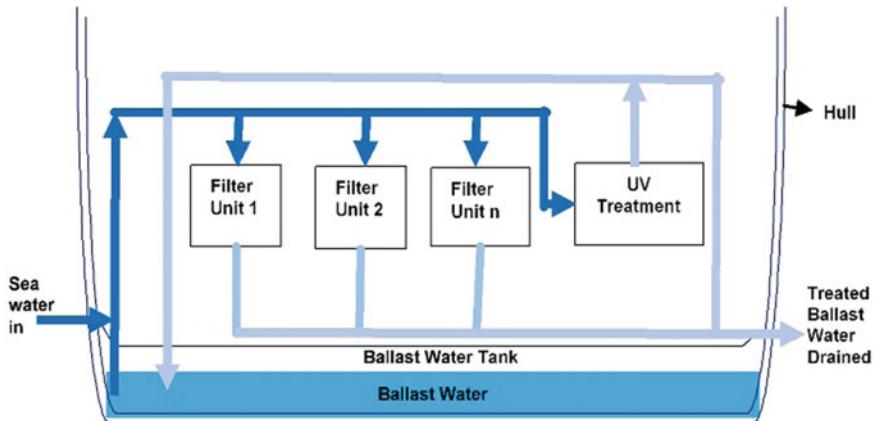


Fig. 2 On-board ship ballast water treatment system

4 IoT Based Smart Ballast Water Quality Compliance

Internet of Things (IoT) Smart Ballast Water (IoTSBW) compliance framework is proposed to enable remote monitoring of water quality of ballast water tank. Different ion sensors [21] are embedded in the ballast water tank to measure the dissolved compounds in water, pH, soil/water temperature and turbidity.

The objective of the proposed architectures given in Figs. 3 and 4 is to gather the ballast water quality parameters from each ship during its loading and unloading of ballast water. The observed values collected are analyzed at a central server or a cloud server located under the control of a statutory body, which can take legal actions against the ship owner in case of non-compliance of standards. The regulation given in BWM convention lays out that every ship is to maintain a record book during every voyage for verification. This paper has attempted to employ an intelligent and automated system using IoT to collect, organize, store and analyze using *big data intelligence* system. GPS module records the location of the ship and the timestamp during which the data are collected. Each ship transmits the data to the central or cloud server using the satellite communication. A mobile application is also a part of this proposal, which connects the officers in a ship to this framework and informs the quality parameters of ballast water sampled recently. Alerting through audio alarm and display messages are part of this application to alert the personnel and be notified immediately when the BWT system malfunctions.

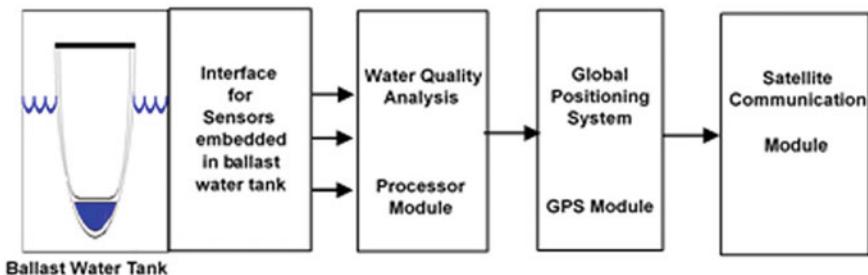


Fig. 3 Proposed architecture for an on-board smart ballast water quality monitoring system installed in a ship

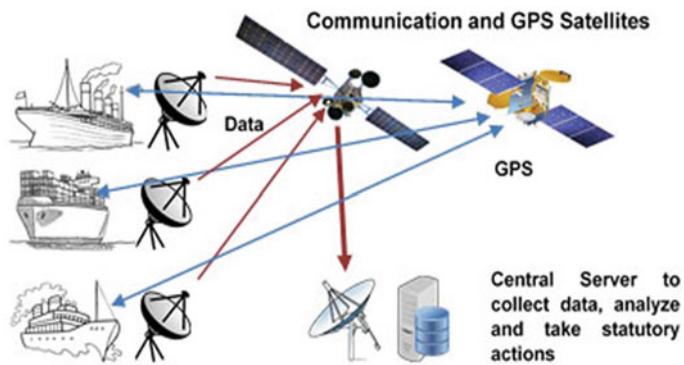


Fig. 4 Proposed framework for smart ballast water quality monitoring system for the statuary agency, IMO

4.1 Model Based on Internet of Things and Big Data Analytics

Figure 4 illustrates the layout of the components involved in transmitting, storing and analyzing the data on treated water quality before draining into the sea. Each ship transmits the data and the GPS to the central server located at the controlling authority. Rather than maintaining a manual record of water treatment schedules and data at the ship level, it is proposed to keep the electronic records and concurrently to send the digital data to the compliant authority as and when the data are captured while executing the sampling plans. The Central server collects and analyses as per policy terms to ship category. Big data analytics and intelligence, modelled by Koga Sadaharu [22] is established in the server, which helps to visualize the summary on each compliance parameter periodically. Query optimization techniques [23] are utilized so as to optimize computing and transmission speed. Trend analysis results are also stored for checking whether sampling are based on data characteristics and characteristic tuples are extracted according to relation compactness [24].

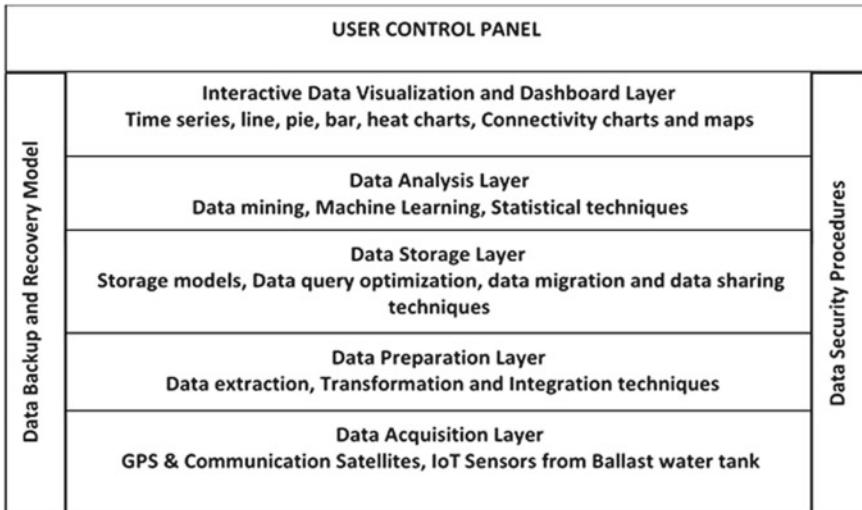


Fig. 5 Big data analytics model

Figure 5 illustrates a layered architecture to collect, organize, store and analyze the big data derived from unstructured data sources such as IoT and satellites. The layers are constituted for (i) data acquisition, (ii) data preparation, (iii) data organization, (iv) data analysis, (v) data visualization, (vi) data security, and (vii) data backup and recovery. Secret keys are used in data sharing and dissemination among the authorized users. Big data, however, faces a stern bottleneck when it applies encrypting techniques for a huge volume of data [25], size of which generally be in terms of petabytes. The real time processing and the data volume are unfavorable to cryptographic encryption due to its heavy computing overheads. A need for a secure but faster encryption technique is increasingly required. Big data analysis applies data analytic techniques such as data mining, deep learning, and statistical techniques to find trends and patterns from real time, unstructured and massive data. Data backup procedure make periodical storage and in case of emergency, the recovery procedure will bring data back alive from the stored.

4.2 *Study of the Proposed Model*

A prototype model is designed with a controller, sensor, gateway, 4G wireless and a GPS interface module. Each node will transmit the following data captured in real time to the server through the gateway where the data get stored for analysis.

- Timestamp
- Geographical location of the node

- Ion measures for ammonium, bromide, calcium, chloride, cupric, fluoride, iodide, lithium, magnesium, nitrate ion (NO_3^- and NO_2^-), perchlorate, potassium, silver and sodium
- pH, turbidity and soil/water temperature
- Volume of water sampled.

The data acquired from the sensors are calibrated and transmitted to the server. The prototype model is tested with the different water samples and the data collected are analyzed to value the quality of the water samples considered. Query optimization, pattern analysis and data security techniques are all tested separately. Non-compliant parameters are displayed to the users on their mobile devices for every sample monitored. The outcomes are found to be consistent and the investigation shows that the model is suitable for remote monitoring of the ballast water quality.

5 Conclusion

Ballast water tank carries aquatic organisms that could cause harmful impact on environment, economy and health when discharged into sea. The BWM convention, 2004 expects the cargo ships strictly adhere to the standards of the ballast water before draining into sea. This paper introduced a remote monitoring model using Internet-of-Things that continuously evaluate the data collected from different samples of ballast water. With this framework, any statutory institute such as IMO could initiate actions against non-compliant ship owners. An IoT technology is an ideal platform to connect the ships through the satellite communication to collect the data from anywhere. Compliance of vessels on BWT cannot easily be arrived without a universal standard for ballast water sampling [26], which is very crucial for handling reliable statuary action.

References

1. IMO: International convention for the control and management of ships' ballast water and sediments. BWM/CONF/36 (2004)
2. Albert, R., Everett, R., Lishman, J., Smith, D.: Availability and efficacy of ballast water treatment technology. EPA's Science Advisory Board Review Paper (2010)
3. Banerji, S., Werschkun, B., Höfer, T.: Assessing the risk of ballast water treatment to human health. *Regul. Toxicol. Pharmacol.* **62**(3), 513–522 (2012). <https://doi.org/10.1016/j.yrtph.2011.11.002>
4. Assessment of the efficacy, availability and environmental impacts of ballast water treatment systems for use in California waters. The California State Legislature (2007)
5. Werschkun, B., Banerji, S.: Emerging risks from ballast water treatment: The run-up to the International ballast water management convention. *Chemosphere* **112**, 256–266 (2014). <https://doi.org/10.1016/j.chemosphere.2014.03.135>

6. David, M.: Vessels and ballast water. In: David, M., Gollasch, S. (eds.) Global Maritime Transport and Ballast Water Management—Issues and Solutions, pp. 13–34. Springer, New York (2015) <https://doi.org/10.1007/978-94-017-9367-4>
7. Briski, E., Gollasch, S., David, M., Linley, R.D., Rajakaruna, H., Bailey, S.A.: Combining ballast water exchange and treatment to maximize prevention of species introductions to freshwater ecosystems. *Environ. Sci. Technol.* **49**, 9566–9573 (2015). <https://doi.org/10.1021/acs.est.5b01795>
8. IMO: Guidelines for approval of ballast water management systems (G8), Resolution MEPC 174(58) (2008)
9. IMO: Procedure for approval of ballast water management systems that make use of active substances (G9). Resolution MEPC 169(57) (2008)
10. IMO: Guidance to ensure safe handling and storage of chemicals and preparations used to treat ballast water and the development of safety procedures for risks to the ship and crew resulting from the treatment process. BWM.2/Circ.20 (2009)
11. IMO: Methodology for information gathering and conduct of work of the GESAMP-BWWG. BWM.2/Circ.13/Rev.1 (2012)
12. Guidelines on ballast water management. Indian Register of Shipping. Report no. IRS-G-ENV-01 (2016)
13. Hansen, L.F., Ølggaard, L.: Order on the management of ballast water and sediments from ships' ballast water tanks. Danish Environmental Protection Agency, Order no. 968 of 24 July 2017
14. Ballast water treatment technologies and current system availability. Lloyd's Reg. Underst. Ballast Water Manage. Ser. (2012)
15. Schillak, L., Heuser, H.P.: Effective new technologies. Federal German Hydrographic and Maritime Agency, Report no. 0014-00007/R0 14/11/2014 (2014)
16. Murphy, K.R., Field, M.P., Waite, T.D., Ruiz, G.M.: Trace elements in ships' ballast water as tracers of mid-ocean exchange. *Sci. Total Environ.* **393**, 11–26 (2008). <https://doi.org/10.1016/j.scitotenv.2007.12.011>
17. Zofia, J., Barańska, M.: Ecological risk assessment of ballast water. In: 1st International Conference Green Cities 2014—Green Logistics for Greener Cities. Procedia—Social and Behavioral Sciences, pp 122–126. Elsevier Ltd, Szczecin, Poland, 19–21 May 2014
18. Bakalar, G., Baggini, M.B.: A verification of a remote monitoring of results of ballast water management system. In: Third International Conference on Traffic and Transport Engineering ICTTE 2016, Belgrade, pp 322–327 (2016)
19. Bakalar, G., Baggini, M.B.: Automated remote method and system for monitoring performance of ballast water treatment system operation on ships. In: 58th International Symposium ELMAR 2016, pp 229–232 (2016)
20. Vorkapića, A., Komarb, I., Jelić Mrčelić, G.: Shipboard ballast water treatment systems on seagoing ships. *Trans. Marit. Sci.* **1**(1), 19–28 (2016). <https://doi.org/10.7225/toms.v05.n01.003>
21. Smart Water Ions Technical Guide: <http://www.libelium.com>
22. Koga, S.: Major challenges and solutions for utilizing big data in the maritime industry. World Maritime University Dissertations (2015). http://commons.wmu.se/all_dissertations/490
23. Thomas, S., Kevin, L.: MapReduce optimization using regulated dynamic prioritization. In: 11th International Joint Conference on Measurement and Modeling of Computer Systems SIGMETRICS'09, pp 299–310. Seattle, WA, USA, 15–19 June 2009
24. Zhang, J., Chen, G., Tang, X.: Extracting representative information to enhance flexible data queries. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(6), 928–941 (2012). <https://doi.org/10.1109/tnnls.2012.2193415>
25. Toshniwal, R., Dastidar, K.G., Nath, A.: Big data security issues and challenges. *Int. J. Innovative Res. Adv. Eng.* **2**(2), 15–20 (2015)
26. Gollasch, S., David, M.: Sampling methodologies and approaches for ballast water management compliance monitoring. *Traffic Transp.* **23**(5), 397–405 (2011). <https://doi.org/10.7307/ptt.v23i5.158>

Analysis of Malignancy Using Enhanced GraphCut-Based Clustering for Diagnosis of Bone Cancer



B. S. Vandana, P. J. Antony and Sathyavathi R. Alva

Abstract Osteosarcoma and Ewing's sarcoma are very common bone tumors, and its biopsy is characterized with spatial distributions of osteoblasts, osteocytes, and osteoclasts. Any abnormal growth found in these three cells can be either cancerous or benign. This paper presents enhanced GraphCut-based clustering framework to ascertain malignancy level in hematoxylin and eosin (H&E)-stained histopathological images. This approach executes iterative GraphCut method to extract foreground objects from biopsy image. Usually, iterative GraphCut needs user interaction to initialize segmentation process. But in enhanced GraphCut method, this initial data is manually generated using standard image processing tools. By doing this, experiment shows that quality of proposed segmentation result is improved. After segmentation of all tissue cells, its categorization is done through color and topological characteristics. Therefore, domain-specific methods such as color-based clustering, mathematical morphology, and active contour are used for feature extraction. This computed features are used to quantify the characteristics of malignancy and classify them as normal, benign, and cancerous using multiclass random forest framework. Proposed method is compared with earlier methods which yields 90% of classification accuracy.

Keywords Histopathology · Bone biopsy · Cancer diagnosis · GraphCut · Malignancy classification

B. S. Vandana (✉)

Dept of CS & E, KVG College of Engineering, Sullia, Karnataka, India
e-mail: vandanamata@yahoo.co.in

P. J. Antony

Dept of CS&E, AJ Institute of Engineering & Technology, Mangalore, Karnataka, India
e-mail: antonyjohn@gmail.com

S. Alva

Department of Pathology, KVG Medical College, Sullia, Karnataka, India

1 Introduction

Cancer is a major cause of death worldwide accounting for 8.8 million deaths in 2015. Primary bone cancers are infrequent, accounting for less than 0.2% of all cancers [1]. As per the American Cancer Society Standardization, 3300 people will be pinpointed with bone cancer. Around 1490 people are expected to lose their life from bone cancer [2]. It is evident from the statistics that bone cancer is spreading across the globe. Therefore, it is pertinent to focus more seriously on research in health sector. Bone tissue is made up of osteoblasts, osteocytes, and osteoclasts. A destructive growth in these three primitive biological cells of mesenchymal can reveal cell differentiation and malignant outgrowth [3].

Cancer diagnosis process has several stages such as clinical examination, radiological examination, and biopsy for histopathology. Research work in radiology reached its zenith, but there is no significant achievement in automation of histopathology. Therefore, the proposed research work emphasizes development of automation tool in histopathology field. There are unique challenges due to the variation in complex tissue structure. By increased bone cancer and advancement in technology, datasets are readily available for study [4]. This study is indebted to bone cancer with the datasets of Ewing's sarcoma and osteosarcoma.

Computer-aided diagnosis (CAD) systems in histopathology field benefits the pathologist to mitigate the workload and for accurate diagnosis. Moreover, lack of dataset and technical expertise in this field remained unattended [5–7].

Increased workload pressure on pathologists and demand for early diagnosis is need of the hour, and it is inevitable. This research exploration in pathology paved the way for collection of literature and dataset to carry out automation work, and it is acknowledged in the review of literature done in my previous work [8].

Research findings witness the development of several algorithms and novel techniques in exploration for segmentation of cell nuclei. These segmentation techniques such as mathematical morphology [9], watershed segmentation [10], probability-based expectation maximization [11], active contour [12], and region growing methods [13] work on pixel, color, and topological feature with its own discrepancies. It is imperative to conclude from the above-noted research work that accuracy in segmentation of tissue cell is remarkable.

GrabCut is considered as one of the image partitioning method since it requires user interaction by specifying window box around the object of interest. The algorithm calculates the color distribution of the target object and the background using model-based clustering method. The max-flow min-cut method is used to separate foreground and background. Misclassified areas are distinguished and adjusted by rerunning the algorithm. Iterative version of this algorithm has been used for foreground segmentation [14].

Features of segmented images are the component to extract information regarding the state of disease. Pleomorphism, mitotic activity, cell counts determine the diagnosis of cancer [15–17]. Measurement of nuclei count is required for diagnosis, but manual counting of nuclei is too hectic which leads to variation in result and it is

evident from the research work done by Fuchs stated 42% variation in classification of nuclei by pathologists [18]. This uncertainty created the necessity for developing automation framework counting tool. Cell identification and classification of digital slide is difficult because of its complicated morphology. Hence, selection of a classifier is important to deal with large dataset [19]. Health sector needs higher-end technology for easy diagnosis.

Organization

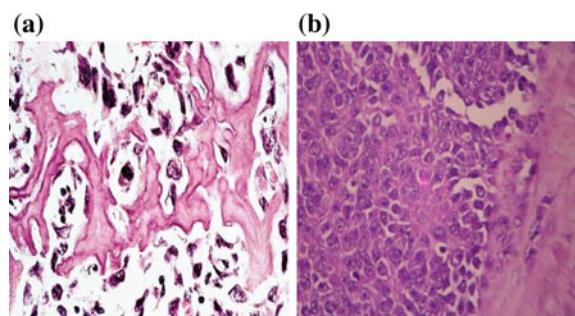
Next part of this paper is organized as follows. Section 2 describes methods to separate malignant structure regions based on enhanced GraphCut method. Feature extraction and classification approach are defined in the same subsection. Accordingly, experimental result's discussion is narrated in Sect. 3. Finally, summary is outlined and future research perspective is highlighted in Sect. 4.

2 Materials and Methods

H&E-stained digital slide of bone biopsy plays pivotal role in this research work. This stained biopsy slide is scanned through a microscope at magnification factor around 40 which converts it into digital form approximately gigabyte size as illustrated in Fig. 1a, b. Pathologist derives diagnosis observations from the biopsy slide. These findings cater the need of a pathologist through simplifying diagnosis process with accurate declaration of disease. Datasets are collected from KVG Medical College & Hospital, Sullia, Kasturba Medical College & Hospital, Mangalore, and Kidwai Institute of Oncology, Bangalore.

It is essential to study the histology of bone, bone-forming cells, osteoid production, mitotic activity, and the common disorder to characterize normal and cancerous image. In osteosarcoma cell, manifold features need to be segmented through automation require the research endeavor. If this research tool succeeds in bone cancer, it can be applicable for all common types of cancer.

Fig. 1 H&E-stained bone slide **a.** Osteosarcoma osteoid with malignant cells **b.** Small-cell Ewing's sarcoma shows large pleiomorphic cells



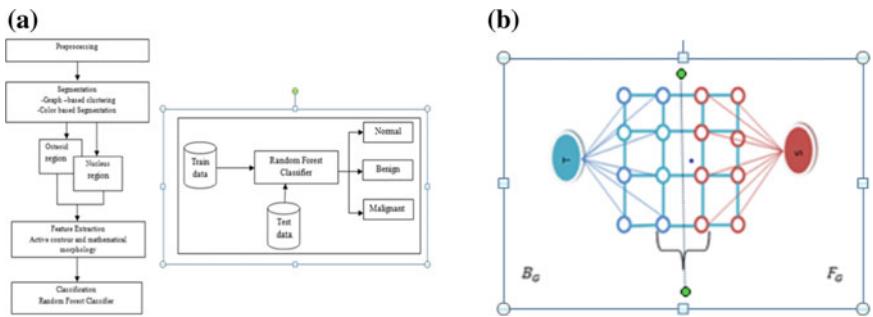


Fig. 2 **a** Framework for bone cancer detection **b** Enhanced GraphCut

Figure 2a shows diagram of automated system which classifies their malignancy state as normal, benign, and cancerous. In this study, we use 140 images of bone biopsy samples and abnormalities which are diagnosed through enhanced GraphCut-based clustering framework.

2.1 Preprocessing

Color is one of the predominant factors for malignant analysis. Therefore, preprocessing is working in RGB color space. Stain artifacts cleared using the technique of Gaussian filters method to maintain sharp intensity for separating tissue cells. Image normalization can be done through histogram modification technique. These preprocessing methods provide high-quality images.

2.2 Image Segmentation

The proposed approach introduces enhanced GraphCut algorithm for the segmentation of different tissue cells in digital bone biopsy. This approach has two phases.

1. GraphCut method to extract foreground objects:

Iterative GraphCut algorithm is used to extract foreground objects in biopsy image. First, user should specify the background pixels by selecting desired foreground object in an image. Algorithm does an initial labeling where all unknown pixels are identified as foreground class F_G and all known pixels are set to background class B_G . Now a Gaussian mixture model (GMM) is defined which learns and creates new pixel distribution. A graph G is made from this pixel cluster in which vertices in the graphs are pixels. Two vertices are selected as seed nodes, namely S and T . Every foreground pixel is connected to S , and every background pixel is connected to T . The edges are weighted based on pixel similarity. The edges are assigned with

minimum cost if there is variance in pixel color. Then a GraphCut algorithm is used to segment the graph based on minimum cost function. Once minimum weight edges are removed, the pixels which are connected to S belong to F_G and those connected to T become B_G . The process is continued until it converges. The process of portioning the region as foreground and background is illustrated in Fig. 2b .

The steps of GraphCut process are given in algorithm 1.

2. Color-based segmentation method:

To learn the malignancy factors of the different types of bone tissue, we run color-based segmentation method which quantizes the pixels of foreground component into two clusters. These clusters correspond to pink and blue regions. Pink color represents osteoid deposition and cytoplasm. Starting with user-specified colors of the objects to be segmented, the algorithm estimates the color distribution of the target objects using a nearest neighbor rule. Each resultant cluster corresponds to blue nuclei and pink osteoid deposition. This result is used in the next level to determine the abnormality in the images.

Algorithm 1: Enhanced GraphCut Method

Input : H&E Image, F_G ,and B_G where F_G is the foreground which is defined as unknown part of the image and B_G known background region. Foreground selection done manually, so that it covers all types of tissue structures.

Output: Segmentation of foreground I_{out} from image.

- 1 Initially label foreground F_G and background B_G pixels depends on the input provided.
 - 2 Train Gaussian Mixture Mode (GMM) based on initial labeling, which is used to model foreground and background pixels as two clusters c_1 and c_2 .
 - 3 Select node S and T from c_1 and c_2 respectively.
 - 4 Construct a graph where c_1 pixels are connected to S and c_2 pixels are connected to T through edges.
 - 5 Assign weight to edge depends on pixel similarity. Low weight is assigned if there is variance in pixel color.
 - 6 Use Min-cut algorithm to remove low cost edges.
 - 7 Refine operation until Min- cut algorithm converges.
-

2.3 Active Contour and Mathematical Morphology-Based Feature Extraction

The proposed approach transforms a segmented tissue component to object domain and pixel domain, respectively. The method uses edge-based active contour method and mathematical morphology to extract lace-like pattern from image component. The specification for active contour method is input image and a mask. For quick and accurate segmentation result, an initial contour nearest to the object shape is specified. All the identified shapes are then labeled. Morphology opening function is performed

to estimate the region between defined lower bound (LB) and upper bound (UB). Finding abnormality in osteoid production alone not defines the state of lesion. The main idea of separating dark blue component is to find out hyperchromatic nuclei. The proposed algorithm is based on finding almost circular object in blue component. The algorithm starts by converting blue components into binary image (B). Finding all connected object in an image and marking each of them with a separate label is called connected region labeling. For analysis of the detected regions, it is required to measure the features like area and perimeter. Using these two measures, the object roundness is calculated as given in Equ. (1).

$$\text{Roundness} = \frac{(\text{perimeter})^2}{4\pi \text{area}} \quad (1)$$

Geometry, morphology, and texture are the low-level features that are used to define abstract concepts such as abnormal mitotic and osteoid production, irregularity of nucleus shape and size. So the relevant features are extracted from segmented image components to build classifier model.

2.4 Random Forest for Malignancy Level Classification

Random Forest is the most popular supervised classification method which works in machine learning for the classification. This serves by building aggregation of decision trees at training time and outputs the class by taking the majority votes of trees. Training set with tissue component labels is used: osteoid deposit, hyperchromatic nuclei, and malignant. Each of this tissue labeled describes specific feature of malignancy. Row in the training set can be a set of attributes to be used as pattern for matching. This study needs three classes, normal, benign, and cancerous, so it requires multiclass. This framework assigns test data as three distinct values (namely 1, 2, and 3). An image classified under class 1 is identified as normal, class 2 as benign, and 3 as cancerous.

3 Result Interpretation

The iterative GraphCut method was tested using a dataset of different bone biopsy images, as shown in Fig. 3. The result completely depends on GraphCut initialization, which makes the segmentation results sensitive. To address this issue, initial selection of object is manually built. Thereby efficiency of the GraphCut method will be improved by not considering unnecessary image elements. For defining the tissue objects, we run the color-based segmentation method on extracted foreground component. The foreground component contains two prominent colored areas such as pink and blue. The result analysis can be seen in Fig. 4a, b.

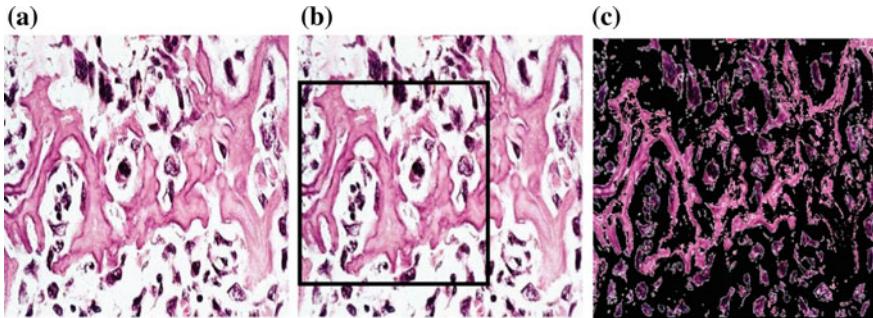
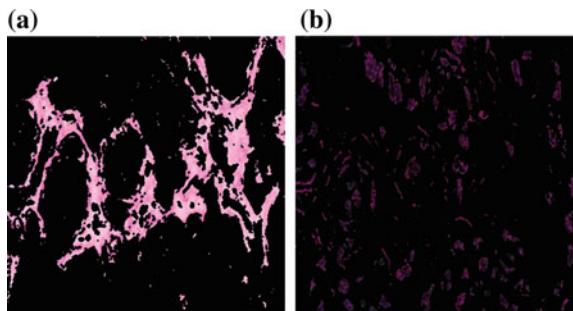


Fig. 3 (a) Input image, (b) selection of initial data, and (c) foreground extraction

Fig. 4 **a** Separation of pink component shows osteoid deposition and **b** separation of dark blue shows hyperchromatic nuclei



3.1 Classification Accuracy

A total of 140 hematoxylin and eosin (H&E)-stained histopathological images comprised osteosarcoma, Ewing's sarcoma, and normal bone (includes benign). A total of 80 images are used for training purpose and remaining images as test set. Training set with tissue component labels is used: osteoid deposit, hyperchromatic nuclei, malignant (pleiomorphic and abnormal mitotic). The corresponding classification accuracy obtained is listed in Table 1. These features are trained through multiclass random forest which gives better performance with 90% accuracy. But earlier studies only stressed upon limited features which led to poor performance.

Table 1 Classification accuracy in the test dataset

	Accuracy measures (%)			
	Recall	Specificity	Precision	Fscore
Normal (8)	87	100	100	93.048
Benign (12)	75	93.7	75	75
Cancerous (40)	95	84.2	92	93.47

Manual Diagnosis of Test Data (Analysis from Experts)

1. Cancerous bone image: There are 40 images selected for the study, out of which 20 images are actually having malignant osteoid features and 20 images show hyperchromatic nuclei pattern. All these images show abnormal mitotic activity also.
2. Normal and benign images: Total 20 images selected for the study, out of which 8 images show normal tissue pattern and 12 images having slightly abnormal tissue pattern stated as benign.

The result of classification model is demonstrated in Fig. 5a–c.

Significance of this study is whatever image used here will be categorized as per the features. Depending upon the output of these, the classifier proposed system defines malignancy level as normal, benign, and malignant. Precision calculates the result relevancy, while recall determines how many truly relevant results are returned. Fscore is the union metric of precision and recall which measures the reconstruction of degree between precision and recall criteria. For normal images, both precision and recall have high scores that show the classifier is returning accurate results. For cancerous images (Class 3) shows high recall relates to a low false negative rate. But slight variation is detected in benign images (Class 2). The reason is some benign images show abnormal pattern of tissue cells due to some infections. Hence, classifier assigns this type of images as class 3 instead of 2.

To investigate the effectiveness of proposed method, we compare with earlier method where malignancy state defined by considering only nuclei count. The experiment shows that our method improves the accuracy by 14% as shown in Table 2.

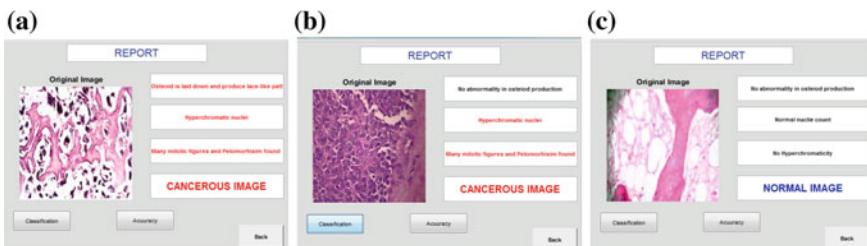


Fig. 5 Automated framework result for **a** osteosarcoma, **b** Ewing's sarcoma, **c** normal image

Table 2 Evaluation of classification result

Diagnosis process	Accuracy (%)
Earlier work depends on nuclei count	76
Proposed work	90

4 Conclusion and Future Work

Clinical approach and manual diagnosis is insufficient to declare disease like cancer. Diversified complex features having different subtypes in bone tissue cells made diagnosis as most challenging one. Hence, automation framework is essential for early and quick accurate diagnosis. Many scholars were tried to explore automation in digital histopathology with limited individual features. This minimal characteristics are not enough for defining the state of disease because of complex tissue cells in bone images. To overcome these obstacles proposed research work combined three categories of feature set in bone cancer for classification of malignancy. Therefore, proposed automated tool framework is commendable, and accuracy of the work is appreciated by pathologists.

Any research work is incomplete without complete dataset and more focused research group. Adequate data sources with dedicated scholars can contribute their best in this field. Future work is particularly intended to increase the strength of data sources and involvement of good number of domain experts to robust techniques.

References

1. Who cancer updates, cancer-world health organization (2017a). <http://www.who.int/cancer>
2. American cancer society cancer facts and figures, American cancer society publication (2017b). <https://www.cancer.org>
3. Heymann, D.: Bone cancer, Primary bone cancers and Bone Metastases. Elsevier Health Sciences (2014)
4. Ghaznavi, F., Evans, A., Madabhushi, A., Feldman, M.: Digital imaging in pathology: whole-slide imaging and beyond. *Ann. Rev. Pathol. Mech. Dis.* **8**, 331–359 (2013)
5. Thiran, J.P., Benoit, M.: Morphological feature extraction for the classification of digital images of cancerous tissues. *IEEE Trans. Biomed. Eng.* **43**(10), 1011–1020 (1996)
6. Weyn, B.: Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. *Cytometry* **33**(1), 32–40 (1998)
7. Einstein, A.J., Wu, H.S., Sanchez, M., Gil, J.: Fractal characterization of chromatin appearance for diagnosis in breast cytology. *J. Pathol* **185**(4), 366–381 (1998)
8. Antony, P.J., Vandana, B.S., Alva, S.R.: Visualisation of tissue cells from bone histopathological images for cancer diagnosis: a review on current status in digital histopathology. *Int. J. Biomed. Eng. Technol.* **24**(2), 133–153 (2017)
9. Veta, M., Pluim, J.P., Van Diest, P.J., Viergever, M.A.: Breast cancer histopathology image analysis: a review. *IEEE Trans. Biomed. Eng.* **61**(5), 1400–1411 (2014)
10. Cloppet, F., Boucher, v.: Segmentation of overlapping/aggregating nuclei cells in biological images. In : Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, IEEE, pp. 1–4 (2008)
11. Fatakdawala, H., Xu, J., Basavanhally, A., Bhanot, G., Ganeshan, S., Feldman, M., Tomaszewski, J.E., Madabhushi, A.: Expectation-maximization-driven geodesic active contour with overlap resolution (emagacor): application to lymphocyte segmentation on breast cancer histopathology. *IEEE Trans. Biomed. Eng.* **57**(7), 1676–1689 (2010)
12. Beevi, S., Nair, M.S., Bindu, G.R.: Automatic segmentation of cell nuclei using krill herd optimization based multi-thresholding and localized active contour model. *Biocybernetics Biomed. Eng.* **36**(4), 584–596 (2016)

13. Zucker, S.W.: Region growing: childhood and adolescence. *Comput. Graph. Image Process.* **5**(3), 382–399 (1976)
14. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM transactions on graphics (TOG), vol. 23, pp. 309–314. ACM (2004)
15. Huang, C.H.: Semi-supervised color decomposition for histopathological images using exclusive component analysis. In: Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on, IEEE, pp. 1–6 (2015)
16. Stierer, M., Rosen, H., Weber, R.: Nuclear pleomorphism, a strong prognostic factor in axillary node-negative small invasive breast cancer. *Breast Cancer Res. Treat.* **20**(2), 109–116 (1991)
17. Dunne, B., Going, J.J.: Scoring nuclear pleomorphism in breast cancer. *Histopathology* **39**(3), 259–265 (2001)
18. Fuchs, T.J., Buhmann, J.M.: Computational pathology: challenges and promises for tissue analysis. *Comput. Med. Imaging Graph.* **35**(7), 515–530 (2011)
19. Wang, M., Zhou, X., Li, F., Huckins, J., King, R.W., Wong, S.T.C.: Novel cell segmentation and online svm for cell cycle phase identification in automated microscopy. *Bioinformatics* **24**(1), 94–101 (2007)

Personalized News Media Extraction and Archival Framework with News Ordering and Localization



Shine K. George, V. P. Jagathy Raj and Santhosh Kumar Gopalan

Abstract Indigenous knowledge is a collective fact generated by local communities all over the world about nature, technology, health, natural resources, etc. The preservation and propagation of this valuable resource can be effectively achieved through media. Presently, information and communication technology plays a vital role in daily lives of the people. The news desk journalist works on the principle of accessing new archives and extracting an idea from the past and reworking it with the present. Yet scrolling through this tremendous amount of information can be hectic. The proposed system of personalized news extraction helps the journalists in identifying the semantically matching news by making use of ontology. Localization and news ordering are the two main features included in the proposed sustainable framework. The localization focuses on extracting news about traditional knowledge. One of the potential elements that determine the news value of a news story is its occurring place. News ordering deals with the listing of news for a query that is specific to the location from where it has been broadcasted. Further studies are required to improve news ordering by discovering how each news event affects the lives of the local community.

Keywords Artificial intelligence · Natural language processing · Ontology · Traditional knowledge · Machine learning

S. K. George (✉) · V. P. Jagathy Raj · S. K. Gopalan
Cochin University of Science and Technology, Cochin, Kerala, India
e-mail: shineucc@gmail.com

V. P. Jagathy Raj
e-mail: jagathy@cusat.ac.in

S. K. Gopalan
e-mail: san@cusat.ac.in

1 Introduction

Visual media plays a mammoth role in updating daily news for everyone in the present scenario. Each piece of news is crafted by connecting the historical and latest scraps of information. News channels archive the past news, related bites, and rushes using computers or tape containing very short descriptions in the form of text for future reference [1].

Most of the precious traditional knowledge (information, innovation, and local practices) from different communities all across the world that preserves a wide range of valuable medicinal plants, herbal formulations, agricultural knowledge, forest products, and handicrafts [2] were not written down as official documents. Local knowledge gave a boost to improve the development of agriculture. Local knowledge if used effectively can mold on innovation and development on the same plane. There is a scope for a wide range of possible innovations based on traditional knowledge and can contribute in reducing poverty in economically backward countries [3–9] and in triggering the ecological restoration.

A centralized news desk that operates as a broadcasting center for multiple news channels would reduce maintenance cost to an extent. The worthy news is what people are keen to know. Geographical location of news happening is one of the measures of news value [10]. Whenever a journalist enquires for news from specific geographical location, they may not get the desired results from the existing system. The proposed news ordering feature provides the journalist with a result based on respective geographical region. It displays those events which are relevant to viewer's physical location. People are more interested to get the news update that are respective to their local area rather than that has happened somewhere else.

The Open Calais ontology developed by Thomson Reuters follows the standardization formed by International Press Telecommunications Council (IPTC). It is specific to the news domain and can analyze and extract the suitable key terms from the news contents and gives the results in the form of social tags, entities with respect to the topic, events, and relations. The proposed framework for news extraction is ontology-based, and the Open Calais is not sufficient to extract the traditional knowledge. So the proposed system uses a local ontology system that takes care of the traditional knowledge. The local ontology and the generic ontology work in a combined fashion where the data is retrieved from both of them for a journalist query based on the requirement. The YouTube-8M dataset [11] is used to evaluate the proposed framework.

This paper is organized as follows: Section 2 summarizes the related works, and the proposed framework is described in Sect. 3. In Sect. 4, experimental results are analyzed, and finally, in Sect. 5, the conclusions and future scope are explained.

2 Related Works

News generation is a continuous and tremendous process in which a number of chores like ordering, storing the whole news contents in a library, news cataloging, and news describing on a fast pace prove to be challenging. Semantic Web technologies can provide a solution to the issues related to news archiving [12].

Time constraints that the journalist faces make him skim through the content for its essence rather than browsing through the whole range of babbling stories. This situation demands a smarter news extraction model leaving behind the conventional models. Semantic modes of news extraction systems are essential to create latest news updates considering the quantity of news and the lack of time availability [13].

To give a personalized flavor to news reports, ontology is the best recommended tool. News extraction is effective if only the machine understands the search query. Ontology has proved its capability of information extraction [14–19] and also has shown the potential in understanding the semantics blended within the news.

3 Proposed News Extraction Framework

The main purpose of this framework is to extract news based on user preference. The proposed framework concentrates on two main functions. First is extracting the news connected to local knowledge using local ontology importing details from a generic ontology if needed. The second feature is how the news has been extracted, ordered based on the proximity of region and listing them on the priority basis. Figure 1 shows the proposed framework.

3.1 *Framework Process Steps*

1. Ontology tags are extracted from the news annotations while archiving news.
2. Keywords along with corresponding news are stored in the database.
3. Using an interface, the journalist enters the query.
4. Tokenization of words from the query is done, and stop words are removed.
5. The query is expanded by extracting equivalent terms from the ontology.
6. The concept terms contained in the query and corresponding keyword terms of each news story stored in the database are matched.
7. The news content that is semantically matched is displayed.

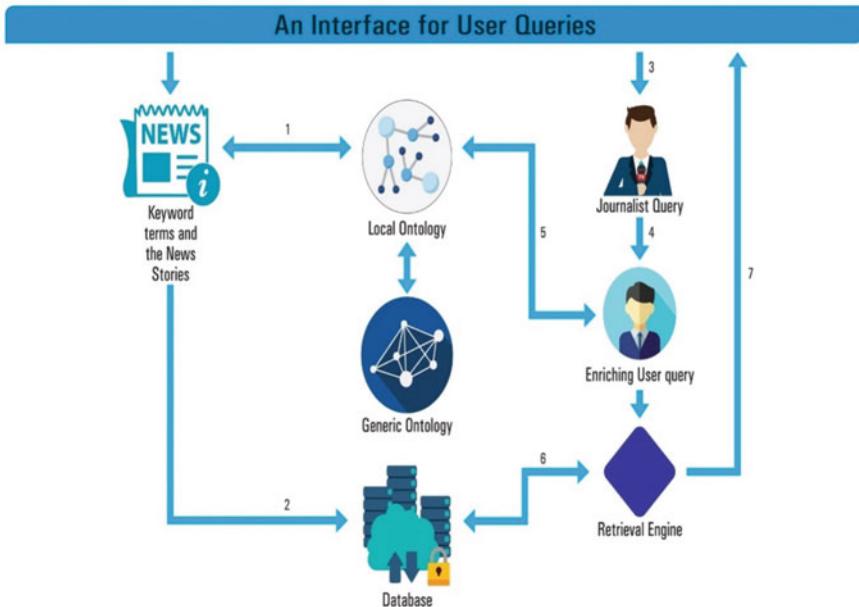


Fig. 1 Ontology-based news extraction framework

4 Evaluation and Results

The application to showcase the proof of concept has been built in Drupal, an effective content management system.

4.1 Local Ontology

A geographical indication (GI) is a representation used on the product that has a particular geographical origin. It should possess a reputation based on that origin. Indigenous knowledge and geographical indications share a common factor in protecting local knowledge to a particular locality. Geographical indication is a tool for traditional knowledge protection [20]. The framework is evaluated using various GIs. ‘Vazhakulam’ is a place in Kerala, India, famous for pineapples. Since the local ontology contains the traditional knowledge, it extracts the news related to ‘vazhakulam pineapple’ for a ‘vazhakulam pineapple’-based query. The news with ‘vazhakulam pineapple’ term will be listed at the top of the search results instead of showing news related to the place ‘Vazhakulam’ and news related to ‘pineapple’.

In order to calculate the semantic similarity between the extracted news stories and the submitted search query, latent semantic analysis (LSA) algorithm is used.

Table 1 Comparison of semantic similarity for sample search queries

Query no.	GI tag	Category	State	Country	Avg. semantic similarity in percentage	
					Generic ontology only	Using local ontology
1	Kachai lemon	Agricultural	Manipur	India	75.1	98
2	Purandar fig	Agricultural	Maharashtra	India	77.6	97.5
3	Malabar pepper	Agricultural	Kerala	India	74.5	98
4	Pokkali rice	Agricultural	Kerala	India	73.8	97.8
5	Vazhakulam Pineapple	Agricultural	Kerala	India	78.1	98.9
6	Aranmula kannadi	Handicraft	Kerala	India	77	98.5
7	Alleppy coir	Handicraft	Kerala	India	74.2	97
8	Bhagalpur silks	Handicraft	Bihar	India	79	97.6
9	Dharwad pedha	Foodstuff	Karnataka	India	71.4	98.7
10	Banglar rasogolla	Foodstuff	West Bengal	India	78.7	98

LSA can effectively analyze the dependency between the key terms and contents given in the text. There are usually many ways to express a given concept. The terms in the document and the literal terms in a user's query may not match. LSA provides a mechanism for the elimination of noise and removal of redundancy. The vector space model is popular in natural language processing. LSA combines vector space model with an analysis model called singular value decomposition (SVD). Thus, group of words' identification of nodes can be related to the vector space that can reflect semantic similarity. The distance among the vectors is then calculated. This calculated distance measures the semantic dependency between texts [21, 22]. Table 1 shows sample results.

Based on the above results, the graph (Fig. 2) is plotted as below that shows the efficiency of local ontology over the generic ontology.

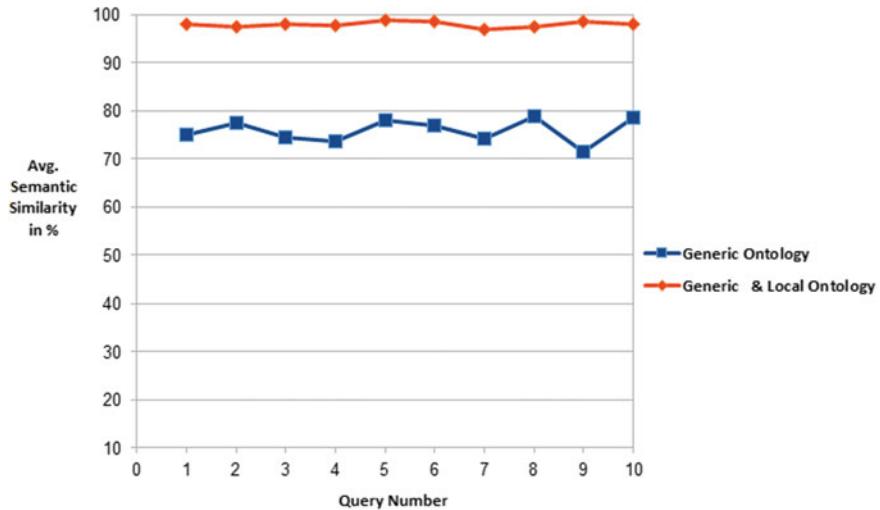


Fig. 2 Localization evaluation result

4.2 News Ordering

In order to get geographical location entities like city, state, provinces, country, and continent, geotags of Open Calais ontology are used in the proposed framework. This way the close geographical locations can be taken for news ordering. News ordering can be set based on district, state, country, and continent. The journalist has to input search query and region. The news ordering framework is tested in YouTube-8M dataset. This dataset includes the collection of videos specific to news broadcasting, current affairs, and other news categories. News videos containing English annotations are taken for the proposed framework. Semantic similarity is measured between search query including region and resultant news stories. The results show that the news stories related to the specific region are listed at the top of the search output. Table 2 summarizes search queries numbered as Q1, Q2, etc., and the evaluation results.

When a journalist searches the news related to Obama to prepare a news story for Asia Edition, the kind of results that the journalist expects is that specific to Asia. The result will be in the following order where the group in ‘bold’ represents the list that appears first and can be in any order and the rest indicates the list that appears last in any order.

Q1, Q3, Q5, Q4, Q6, Q2

Similarly, if a journalist is searching for news related to Obama to prepare a news story for US Edition, the proposed framework produces the results below. The group in ‘bold’ represents the list that appears first and can be in any order and the rest indicates the list that appears last in any order.

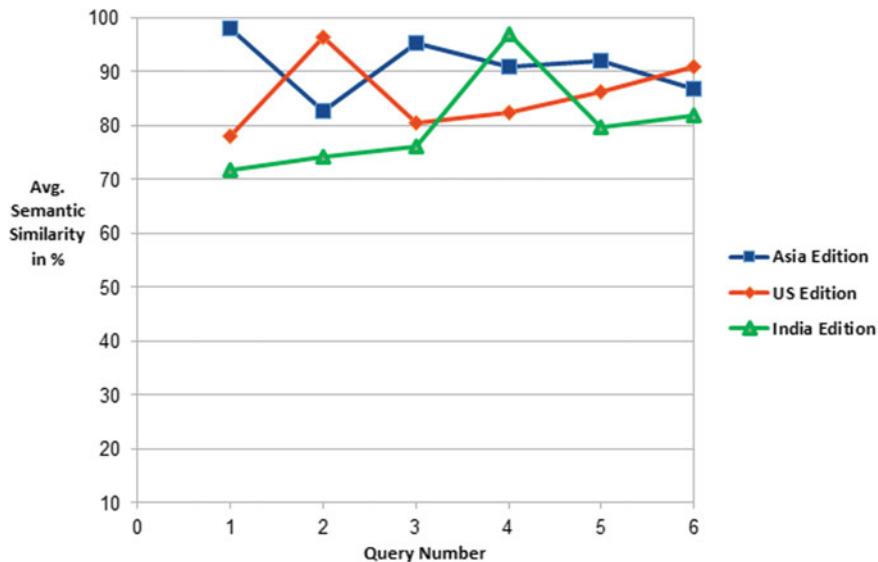


Fig. 3 Comparison of news ordering results on Asia Edition, US Edition, and Indian Edition

Q2, Q6, Q5, Q4, Q3, Q1

For Indian Edition, the results are shown below. The ‘bold’ represents the list that appears first and can be in any order and the rest indicates the list that appears last in any order.

Q4, Q6, Q5, Q2, Q3, Q1

Figure 3 represents a graph which is plotted based on news ordering results of Asia Edition, US Edition, and Indian Edition based on the queries.

Table 2 News ordering results

Query no.	Query	Semantic similarity in percentage		
		Asia	US	India
Q1	Obama about Israel	98	78.2	71.9
Q2	Obama campaign in New York	82.7	96.4	74.2
Q3	Syria war and Obama	95.4	80.5	76.2
Q4	Obama’s visit to Agra	90.9	82.5	97.1
Q5	Obama and Iraq	92.0	86.3	79.7
Q6	Bush meets Obama in Florida	86.7	90.9	81.8

5 Conclusion

In this work, we presented a personalized news extraction framework comprising of two features, namely localization and news ordering. One of the limitations of generic ontology is that it cannot capture indigenous knowledge terms. News focused on traditional knowledge can be extracted using local ontology in the proposed framework. News ordering feature helps to extract the news considering its closeness to the geographical location where the news stories are broadcasted. From the evaluation result, it is clear that the proposed framework delivers better extraction of news and provides the journalist with a better personalization experience.

It is really a hectic task to develop a comprehensive ontology for traditional knowledge. Ontology's completeness improves the news extraction effectiveness. Extracting local context or angle of a news story and discovering how events that occur will affect local community may help improve the prioritization or ordering of news items based on user needs. This can be considered as a scope for future enhancement.

References

- Shine, G.K., Jagathy Raj, V.P., Santhosh Kumar, G.: Ontology based framework for news extraction in visual media. In: Proceedings of IEEE International Conference on Data Science & Engineering (ICDSE), pp. 220–222 (2012)
- Baines, J., Taylor, N., Buckenham, B., Douglas, J., McClintock, W.: The importance of local knowledge. In: 20th Annual Meeting of the International Association for Impact Assessment, Hong Kong (2000)
- Nwokeabia, H.: Why Industrial Revolution missed Africa: A Traditional Knowledge Perspective. Economic Commission for Africa, Addis Ababa (2001)
- Hooli, L.J., Jussi, J.S.: Building an innovation system and indigenous knowledge in Namibia. *Af. J. Sci. Technol. Innovation Dev.* **10**(2), 183–196 (2018)
- Chinsembu, K., Cheikhysussef, A., Mumbengegwi, D., Kazembe, L.: Indigenous Knowledge of Namibia (2015). ISBN:978-99916-42-05-5
- Jauhainen, J.S., Hooli, L.: Indigenous knowledge and developing countries innovation systems: the case of Namibia. *Int. J. Innovation Stud.* **1**(1), 89–106 (2017)
- Materer, S., Valdivia, C., Gilles, J.L.: Indigenous knowledge systems: characteristics and importance to climatic uncertainty. University of Missouri Columbia, Department of Agricultural Economics, Working papers (2002)
- Uprety, Y., Asselin, H., Bergeron, Y., Doyon, B.: Contribution of traditional knowledge to ecological restoration: practices and applications. *Ecoscience* **19**(3), 225–237 (2012)
- Singh, Y., Jasmine Kaur, B., Onial, M., Mathur, V.: Traditional knowledge systems in india for biodiversity conservation. *Indian J. Tradit. Knowl.* **15**, 304–312 (2016)
- Rasyid, A.B.D., Siagian, H.F.: News value and reality: a case study. *Malays. J. Commun.* **33**(3), 55–69 (2017)
- Abu-El-Haija, S., Kothari, N., Lee, J., Natsev, P., Toderici, G., Varadarajan, B., Vijayanarasimhan, S.: YouTube-8M: A large-scale video classification benchmark. arXiv preprint, [arXiv:1609.08675](https://arxiv.org/abs/1609.08675) (2016)
- Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Sci. Am.* 29–37 (2001)
- Wimalasuriya, D., Dejing, D.: Ontology-based information extraction: an introduction and a survey of current approaches. *J. Inf. Sci.* **36**, 306–323 (2010)

14. Wu, F., Weld, D.S., Daniel, S.: Autonomously semantifying wikipedia. In: Proceedings of the 16th ACM Conference on Information and Knowledge Management, pp. 41–50, ACM, Portugal (2007)
15. Kietz, J., Maedche, A., Volz, R.: A method for semi-automatic ontology acquisition from a corporate intranet. In: EKAW-2000 Workshop on Ontologies and Text, pp. 2–6, Juan-Les-Pins (2000)
16. Cimiano, P., Handschuh, S., Staab, S.: Towards the self annotating web. In: Proceedings of the 13th International Conference on World Wide Web, WWW 2004, pp. 462–471, USA (2004)
17. Maynard, D., Peters, W., Li, Y.: Metrics for evaluation of ontology-based information extraction. In: Proceedings of the WWW 2006 Workshop on Evaluation of Ontologies for the Web, ACM, New York (2006)
18. Apurva, D., Gotmare, P.: Semantics based document clustering. *Int. J. Sci. Res. Comput. Sci. Eng.* **5**(4), 25–30 (2017)
19. Chahbar, M., Elhore, A., Askane, Y.: PERO2: machine teaching based on a normalized ontological knowledge base. *Int. J. Sci. Res. Comput. Sci. Eng.* **5**(5), 63–74 (2017)
20. Blakeney, M.: Protection of traditional knowledge by geographical indications. *Int. J. Intell. Prop. Manage.* **3**(4), 357 (2009)
21. Rozeva, A., Zerkova, S.: Assessing semantic similarity of texts—methods and algorithms. In: Proceedings of the 43rd International Conference Applications of Mathematics in Engineering and Economics AIP Conference Proceedings 1910, 060012 (2017)
22. Boling, C., Das, K.: Semantic similarity of documents using latent semantic analysis. In: Proceedings of the National Conference On Undergraduate Research (NCUR) University of Kentucky, Lexington, KY 3–5 Apr 2014

Analysis of Transfer Characteristics of Junctionless GaAs-Nanotube MOSFET with Hafnium Oxide Dielectric



Raj Kumar and Arvind Kumar

Abstract As the feature size of device moving below sub-20-nm regime, it is becoming very challenging to create very sharp junction at source/drain end. Junctionless device has potential to eliminate necessity of creating ultrasharp junctions. Nanotube transistors have been declared superior over nanowire transistor attributed to the presence of two gates. This paper presents the structure of junctionless GaAs-nanotube MOSFET with hafnium oxide as oxide material. The drive current of GaAs-based channel is far more than silicon-based channel of transistor. The junctionless GaAs-nanotube MOSFET has been simulated using VisualTCAD. Junctionless GaAs-nanotube MOSFET with hafnium oxide (HfO_2) dielectric has shown tremendous response over silicon oxide dielectric-based MOSFET in terms of leakage and short-channel effects. The effect of varying design parameters such as channel thickness, oxide thickness and drain voltage on I-V characteristics of device has been studied. Junctionless GaAs-nanotube MOSFET with HfO_2 as oxide and appropriate design parameters can be considered as promising candidate for future CMOS technology.

Keywords Gate all around (GAA) · Nanowire · Nanotube · MOSFET · Junctionless (JL)

1 Introduction

The continuous shrinking of planar MOS transistors has been obstructed due to severe short-channel effects (SCEs) in ultrasmall devices. To continue to stick with downscaling process of devices, researchers came up with some non-conventional CMOS devices such as single- and multiple-gate silicon-on-insulator (SOI) MOSFETs, nanowire and nanotube transistors. However, transistors even with non-conventional

R. Kumar (✉) · A. Kumar

University Institute of Engineering and Technology, Panjab University, Chandigarh, India
e-mail: rajeetkumar77@gmail.com

A. Kumar
e-mail: arvind_uiet@yahoo.co.in

structure below 20 nm need to have ultrasharp doping profile. To fabricate extremely sharp junctions in conventional MOS transistors become huge challenge for future CMOS technology to meet the performance demand for low power consumption and standby power applications [1, 2]. Junctionless transistor (JLT) with uniform heavy doping in source, channel and drain regions eliminates the fabrication of junctions [3, 4]. JLT needs to be fully depleted before it gets ON. JLT works in partial depletion mode unlike the inversion-based transistor. Unlike conventional MOSFET, bulk conduction takes place in junctionless device, which leads to allow shrinking of feature size with improved immunity to short-channel effects and low power dissipation. The electrical performance of nanoscale junctionless MOSFETs declares it as prominent candidate for future CMOS technology [5, 6]. Silicon is no longer a reliable material for CMOS due to its electrical limits. Apart from that, III-V semiconductor materials have been presented as suitable alternate to replace silicon-based transistor for having enhanced device's performance in terms of mobility and driving current [7, 8].

Jin et al. [9], Chiang [10] and Li et al. [11] had presented model of cylindrical junctionless gate-all-around nanowire transistor for threshold voltage and subthreshold characteristics. The design parameters such as thickness of nanowire and oxide caused to affect threshold voltage and subthreshold characteristics of device. Recently, the nanotube transistors have been declared superior to nanowire transistors attributed to two gates (inner and outer) over the channel leading to provide better SCEs immunity and volume inversion. A. Kumar et al. enquired the model for subthreshold characteristics and threshold voltage of Si-nanotube ultrathin double gate-all-around (DGAA) MOSFETs [12, 13]. The thickness of the tube was considered as foremost parameter to control carrier transport properties.

In this paper, GaAs-nanotube n-MOSFET with hafnium oxide HfO₂ has been proposed, designed and simulated using VisualTCAD, a 3D device simulator by Cogenda Pvt. Ltd [14]. The effects of varying design parameters such as oxide thickness, channel thickness and drain bias have been analysed. GaAs-nanotube MOSFET can be further investigated and studied for CMOS circuit's applications.

2 Device Design and Structure

The 3D structural view of the nanotube MOSFET consists of tubular channel with inner gate (core) and outer gate (shell) as shown in Fig. 1. Figure 2 depicts the circular cross-sectional view of device presenting different regions of MOSFET. The inner gate (core) is enclosed by oxide layer which provides additional electrostatics control over the channel region. The device is defined using cylindrical coordinate system and 2D cross-sectional view as shown in Fig. 3.

The device exhibits uniform doping throughout source, channel and drain. Device parameters are used in simulation set-up as described in Table 1. Both inner and outer gates are interconnected to provide same work function and are biased with gate voltage (V_{GS}).

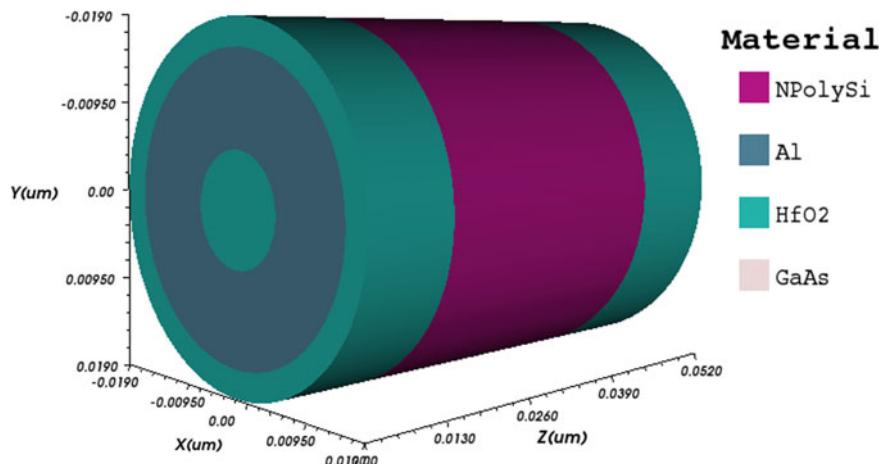


Fig. 1 3D structural view of junctionless GaAs-nanotube MOSFET

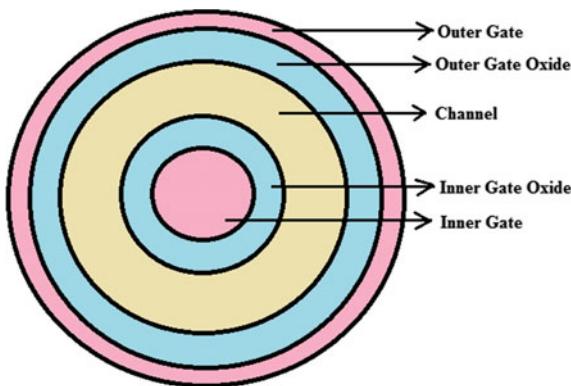


Fig. 2 Circular cross-sectional view of junctionless GaAs-nanotube MOSFET

Table 1 Design parameter for simulation

Design parameter	Values
Gate length (L_g)	30 nm
Channel thickness (t_{GaAs})	5–10 nm
Oxide thickness (t_{ox})	1–3 nm
Core radius (t_c)	5 nm
Oxide material	HfO ₂
Metal work function (Φ)	4.7
Source/drain/channel doping N_D	$1 \times 10^{19} \text{ cm}^{-3}$
Source/drain length	7 nm

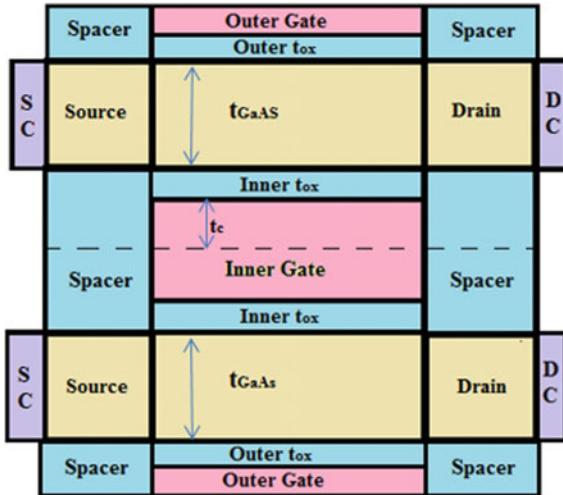


Fig. 3 2D cross-sectional view of junctionless nanotube MOSFET

3 Results and Discussion

This section presents numerical simulation results as I-V transfer characteristics of junctionless GaAs-nanotube MOSFETs by varying the design parameters of device are presented. Visual TCAD, a 3D device simulator by Cogenda Pvt. Ltd., was used to perform simulations. The numerical simulation has been done by incorporating drift diffusion and density gradient, and analytical mobility models are considered for the simulation [14].

Comparison of $\log(I_{ds})$ versus V_{gs} characteristics of JLT GaAs-nanotube FET between SiO_2 and HfO_2 is depicted in Fig. 4; it has been observed from simulation results that drive current increases significantly and leakage current decreases with huge value on replacing silicon oxide (SiO_2) with high-k HfO_2 as gate oxide.

Figure 5 shows the variation in drain current corresponding to gate voltage for different gate oxide thickness (T_{ox}) while keeping other parameter constant. It is observed that leakage current increases abundantly as thickness of oxide increases from 1 to 3 nm attributed to weak electrostatic control over the channel.

As the thickness of nanotube is considered as foremost parameter to control carrier transport properties, it is very imperative to investigate the performance of device by varying nanotube's thickness. It has been investigated from simulation results that subthreshold leakage decreases manifold as the channel thickness reduces from 10 to 5 nm as shown in Fig. 6. This happens due to very strong 2D quantization of the carriers in channel [12]. However, the ON-current is decreasing as reduces channel thickness attributed to strong effect of electric field on conduction mechanism leads to hot carrier's effects. Thus, subthreshold leakage can be reduced by reducing channel thickness.

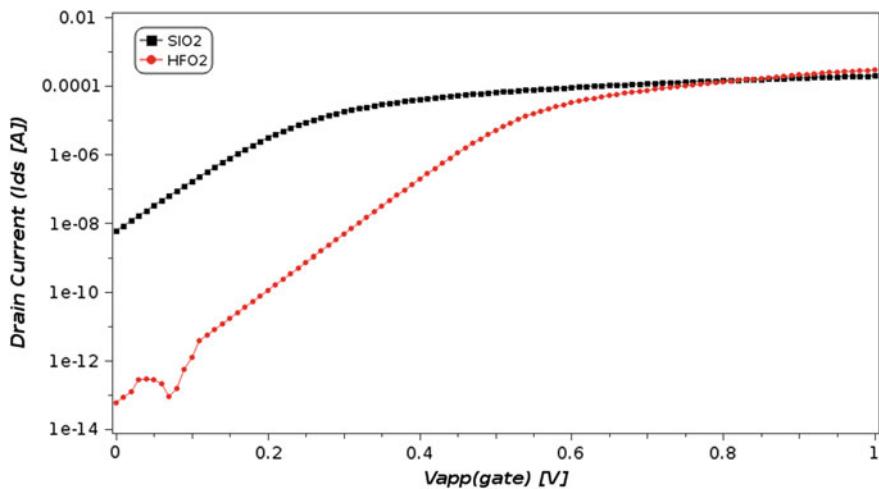


Fig. 4 Comparison of $\log (I_{ds})$ versus V_{gs} characteristics of JLT GaAs-nanotube FET between SiO_2 and HfO_2 with $T_{\text{ox}} = 2 \text{ nm}$, $T_{\text{GaAs}} = 10 \text{ nm}$ and $V_{\text{ds}} = 0.05$

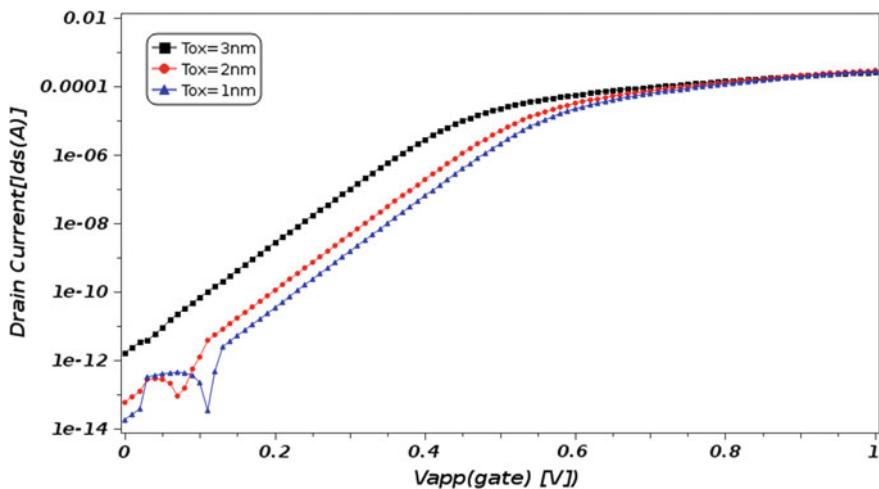


Fig. 5 $\log I_{ds}$ versus V_{gs} characteristic curves of JLT GaAs-nanotube FET for different gate oxide thickness (T_{ox}) with $T_{\text{GaAs}} = 10 \text{ nm}$ and $V_{\text{ds}} = 0.05$

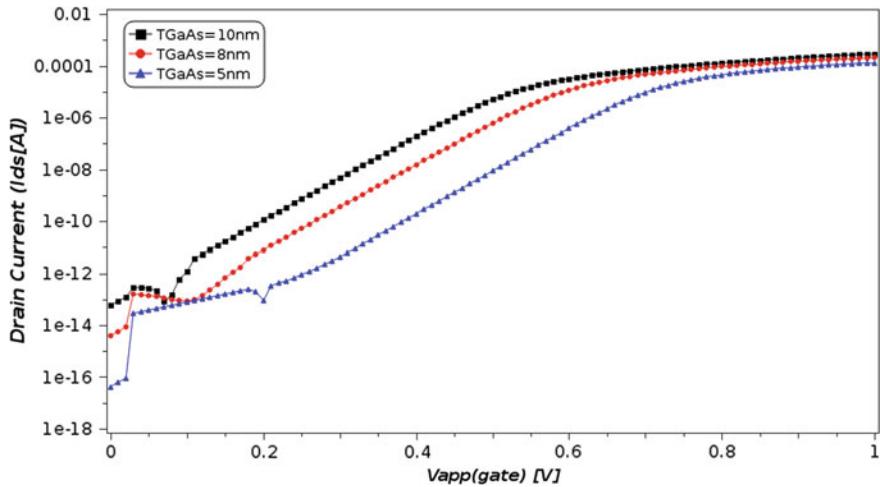


Fig. 6 Log I_{ds} versus V_{gs} characteristics of JLT GaAs-nanotube FET for different channel thickness (T_{GaAs}) with $T_{ox} = 2$ nm and $V_{ds} = 0.05$

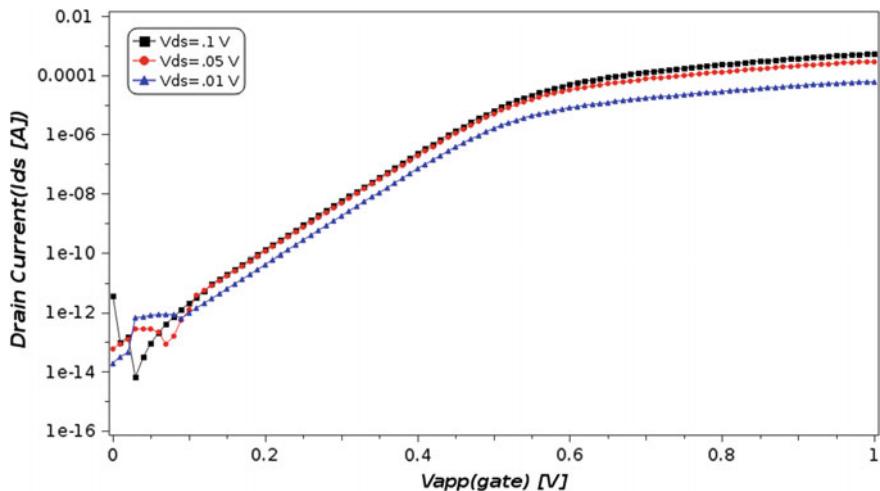


Fig. 7 Comparison of log I_{ds} versus V_{gs} characteristic curves of JLT GaAs-nanotube FET for different drain biasing (V_{ds}) with $T_{ox} = 2$ nm and $T_{GaAs} = 10$ nm

Comparison of log I_{ds} versus V_{gs} characteristic curves of JLT GaAs-nanotube FET for different drain biasing is depicted in Fig. 7. It is found from the figure that with increasing the drain bias voltage, subthreshold current increases tremendously attributed to the DIBL effect. This leakage shows the band-to-band tunnelling (BTBT) in off-state, responsible for static power dissipation. Thus, BTBT is the biggest challenge for low standby power (LSTP) applications [14].

4 Conclusion

In this paper, the 3D simulation of junctionless GaAs-nanotube n-MOSFET was performed. It is observed that III-V compound material (GaAs)-based junctionless nanotube transistor is providing high drive current. Hafnium oxide (HfO_2) is best substitute to replace silicon oxide (SiO_2) as gate oxide to reduce the leakage current of device at OFF-state. Subthreshold current decreases with the decrease in oxide thickness and channel thickness, whereas subthreshold leakage increases with increase in drain bias due to DIBL. Junctionless GaAs-nanotube n-MOSFET can be further investigated for future CMOS applications.

References

1. Kim, Y.B.: Challenges for nanoscale MOSFETs and emerging nanoelectronics. *Trans. Electr. Electr. Mate* **11**(3), 93–105 (2010)
2. International technology roadmap for semiconductors (ITRS) (2015)
3. Colinge, J.P., Lee, C.W., Afzalian, A., Akhavan, N.D., Yan, R., Ferain, I., Razavi, P., et al.: Nanowire transistors without junctions. *Nature Nanotechnol* **5**(3), 225–229 (2010)
4. Colinge, J.P., Lee, C.W., Akhavan, N.D., Yan, R., Ferain, I., Razavi, P., Kranti, A., Yu, R.: Junctionless transistors: physics and properties. *Semiconductor-On-Insulator Materials for Nanoelectronics Applications*, pp. 187–200. Springer, Berlin (2011)
5. Lee, C.W., Ferain, I., Afzalian, A., Yan, R., Akhavan, N.D., Razavi, P., Colinge, J.P.: Performance estimation of junctionless multigate transistors. *Solid-State Electr.* **54**(2), 97–103 (2010)
6. Wang, Y., Shan, C., Dou, Z., Wang, L., Cao, F.: Improved performance of nanoscale junctionless transistor based on gate engineering approach. *Microelectron. Reliab.* **55**(2), 318–325 (2015)
7. Song, Y., Zhang, C., Dowdy, R., Chabak, K., Mohseni, P.K., Choi, W., Li, X.: III-V junctionless gate-all-around nanowire MOSFETs for high linearity low power applications. *IEEE Electr. Device Lett.* **35**(3), 324–326 (2014)
8. Razavi, P., Fagas, G.: Electrical performance of III-V gate-all-around nanowire transistors. *Appl. Phys. Lett.* **103**(6), 063506–063508 (2013)
9. Jin, X., Liu, X., Lee, J.H., Lee, J.H.: Modeling of subthreshold characteristics of short channel junctionless cylindrical surrounding-gate nanowire metal–oxide–silicon field effect transistors. *Phy. Scripta* **89**(1), 015804 (2013)
10. Chiang, T.K.: A new quasi-2-D threshold voltage model for short-channel junctionless cylindrical surrounding gate (JLCSG) MOSFETs. *IEEE Trans. Electr. Devices* **59**(11), 3127–3129 (2012)
11. Li, C., Zhuang, Y., Di, S., Han, R.: Subthreshold behavior models for nanoscale short-channel junctionless cylindrical surrounding-gate MOSFETs. *IEEE Trans. Electr. Devices* **60**(11), 3655–3662 (2013)
12. Tiwari, P.K., Samoju, V.R., Sunkara, T., Dubey, S., Jit, S.: Analytical modeling of threshold voltage for symmetrical silicon nano-tube field-effect-transistors (Si-NT FETs). *J. Comput. Electr.* **15**(2), 516–524 (2016)
13. Kumar, A., Bhushan, S., Tiwari, P.K.: A threshold voltage model of silicon-nanotube-based ultrathin double gate-all-around (DGAA) MOSFETs incorporating quantum confinement effects. *IEEE Trans. Nanotechnol.* **16**(5), 868–875 (2017)
14. Reference manual: Genius, 3-D Device Simulator, Version 1.9.2-3, Cogenda Pvt. Ltd, Singapore (2017)

Signal Integrity Analysis for Diameter-Dependent Mixed Carbon Nanotube Bundle Interconnects



Vijay Rao Kumbhare, Punya Prasanna Paltani
and Manoj Kumar Majumder

Abstract Recent VLSI technology considers the structure and modeling approach of mixed carbon nanotube bundle (MCB) interconnects. Initially, this research paper models different MCB structures by considering diameter-dependent modeling of interconnect parasitics. Finally, the unique models are used to study the effect of signal integrity using multi-line bus architecture. Crosstalk induced delay is analyzed by various dimensions of MCB using lengths which vary from 100 to 500 μm of global interconnects. Using industry standard HSPICE circuit simulations, it is observed that the overall improvement in crosstalk induced delay is 6.8% for novel MCB structure having single- and multi-walled CNTs (SWCNT and MWCNT) as compared to MCB containing SW- and DWCNT.

Keywords Mixed CNT bundle (MCB) · Carbon nanotube (CNT) · Signal integrity · Propagation delay · Interconnect · Crosstalk

1 Introduction

Carbon nanotubes (CNTs) have stimulated huge research motivation to many researchers for their applicability as VLSI interconnect and are proved to be a potential candidate in future integrated circuits [1]. CNT exhibits extremely high electrical properties due to its unique band structure of graphene having the zero effective mass of holes and electrons [2]. Based on the current scenario, conventional materials for interconnect such as aluminum (Al) and copper (Cu) exhibit skin effect, electro migration [3, 4], and increased resistive parasitic problem at nanoscale region [5–7].

V. R. Kumbhare · P. P. Paltani · M. K. Majumder (✉)
Dr. S. P. Mukherjee International Institute of Information Technology, Naya Raipur, India
e-mail: manojk@iiitnr.edu.in

V. R. Kumbhare
e-mail: vijay@iiitnr.edu.in

P. P. Paltani
e-mail: punya@iiitnr.edu.in

Dimensionally, the CNT can be created by the ratio of length to the diameter of up to the range of 132,000,000:1 [8], that is considerably larger than other existing materials. CNTs represent several properties such as long ballistic transport phenomenon, large current carrying capability [9, 10], high thermal conductivity [11], and mechanical strength [12, 13].

Graphene sheet is rolled up into hollow cylindrical structure to form CNTs, known as allotropes of carbon [8]. According to rolling up graphene sheet, CNTs can be categorized as single-walled CNT (SWCNT), double-walled CNT (DWCNT), and multi-walled CNTs (MWCNTs). CNTs' structure basically depends on chiral indices that can be identified based on rolling direction of graphene sheet. The chiral indices (n, m) decide the behavior of CNTs as either zigzag or an armchair structure. If chiral indices defined by $n = m$, it will act as armchair CNTs and behave as zigzag if n or $m = 0$ [9]. CNTs can act as metallic and semiconductor depending upon chiral indices. Metallic behavior is always present in armchair CNTs, whereas zigzag is either metallic or semiconducting in nature [9, 10] depending on the value of integer i which satisfies, $n - m = 3i$.

Hypothetically, CNT bundles may contain only one type of CNTs like SWCNTs or MWCNTs or DWCNTs. However, SWCNTs and MWCNTs both can be observed in a realistic CNT bundle that can be randomly arranged in the bundle. MCBs can have many combinations such as SWCNTs and MWCNTs or SWCNTs and DWCNTs or MWCNTs having different diameters, etc. Recent research shows that randomly arranged MCBs can be arranged to a specific requirement by a method known as atomic force manipulation technique [14].

The paper is organized as follows: Sect. 1 presents the introduction of the recent research scenario, and Sect. 2 describes the geometrical structure and equivalent *RLC* circuit model of mixed CNT bundle interconnect. The signal integrity for different interconnect lengths at various technology nodes is presented in Sect. 3. Finally, a brief summarization of this paper is presented in Sect. 4.

2 Equivalent *RLC* Model

This section presents a brief description of the equivalent model for proposed MCB that is used to achieve the crosstalk induced delay. Transmission line model for lumped *RLC* network is considered accounting necessary parasitic effect associated with the bundle.

2.1 Geometry of Mixed CNT Bundle

The MCB is obtained when two or more types of CNTs are accommodated simultaneously in a bundle. The structures that are studied are of specific arrangement and are shown in Fig. 1. This paper models the MCB by considering three levels shown in

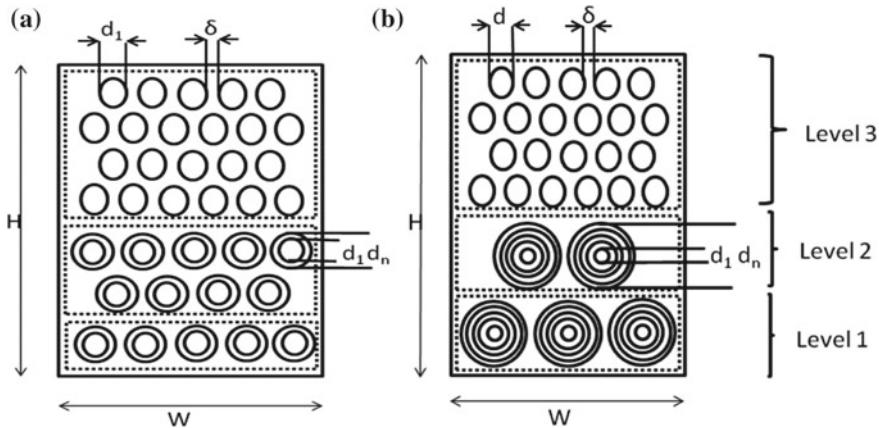


Fig. 1 Structure of mixed CNT bundle **a** MCB-I and **b** MCB-II

Fig. 1 as follows: (i) Level 1 comprises the CNTs in the lower half of MCB structure facing the ground (N_A). (ii) Level 2 comprises the CNTs left in the lower half (N_B), i.e, other than Level 1, and (iii) Level 3 comprises the CNTs in the upper half part of MCB structure (N_C).

Figure 1a presents the MCB-I structure in which the upper half is filled with SWCNTs while the lower one with DWCNTs, whereas Fig. 1b shows the structure of MCB-II in which lower half DWCNTs are replaced by MWCNTs. In Fig. 1, SWCNTs' diameter denoted as d , δ is the gap between the SWCNTs or DWCNTs or MWCNTs, d_1 is the inner shell diameter of DWCNTs or MWCNTs, and d_n is the outer shell diameter of DWCNTs or MWCNTs. The height and width of the CNT bundle are depicted by H and W , respectively. From geometry, it can be revealed that half of the bundle is occupied by SWCNTs and the other half is by DWCNTs or MWCNTs. Figure 2 shows the bundle architecture of MCB that considers different types of coupling capacitance between two levels and two bundles.

2.2 Equivalent RLC Model of Proposed MCB

The proposed model based on equivalent *RLC* for MCB is developed using transmission line theory. An *RLC* parameters are obtained for MCB-I and MCB-II structures that are employed according to Fig. 3.

MWCNTs composed of numerous concentric shells are arranged in parallel. It may be similar to SWCNT bundle consisting of a large number of parallelly arranged nanotubes, but in practical it is significantly different from SWCNT bundle. The conducting channel of MWCNT depends on number of shells, and it will increase for more number of shells. Hence, the combination of each shell parameters is not as simple as in SWCNT bundle scenario. All the circuit parametric variation can

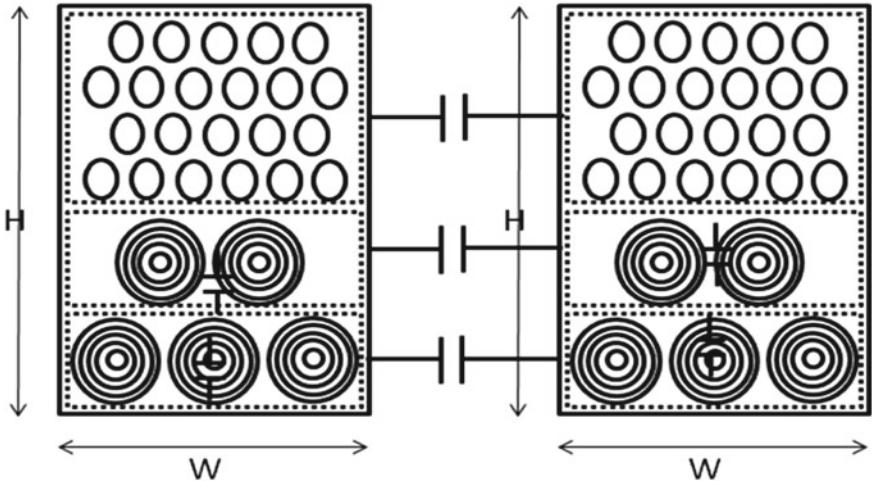


Fig. 2 Bundle architecture showing different capacitance effect

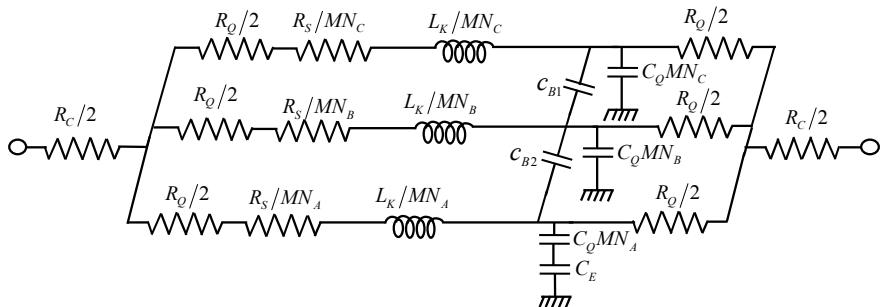


Fig. 3 Equivalent distributed *RLC* circuit model of MCB

be observed for different shell in MWCNTs that cannot be assumed to be equal for SWCNT bundles, which induced coupling capacitance between shells. The small separation between adjacent shells retains a large value of coupling capacitance that is a type of electrostatic capacitance. Modeling of *RLC* transmission line requires the analysis of diameter-dependent interconnect parasitics that have a significant effect on crosstalk delay. Interconnect parasitics include contact and quantum resistances, self- and kinetic inductances, electrostatic, quantum, and coupling capacitances. The parasitic further can be of various forms that need to be modeled. All different types of parasitic values are considered in following subsections.

a. No. of conducting channels

The number of channels in SWCNT is 2 [2], whereas in DWCNT or MWCNT basically depends on the diameter of CNTs [15] and given as

$$N_{\text{shell}} \approx 0.0612 \text{ nm}^{-1} \cdot D + 0.425, D > 3 \text{ nm} \quad (1)$$

where D stands for shell diameter. If it is greater than 3 nm, we can use this expression, and error introduce will be 15% [15]. In MWCNT, the below expression shows the total number of shells denoted by p ,

$$p = 1 + \text{Inter}\left[\frac{(D_n - D_1/2)}{2d}\right] \quad (2)$$

Therefore, the total number of channel is defined by the summation of conducting channels with all the diameters, i.e,

$$N_i = 0.0612 \text{ nm}^{-1} \cdot D_i + 0.425 \quad (3)$$

b. Interconnect parasitics

The resistance of a shell is categorized as scattering-induced resistance (R_S), quantum resistance (R_Q), and imperfect metal–nanotube contact resistance (R_{mc}). If the mean free path (mfp) is greater than tube length, the effect of scattering resistance (R_S) can be observed. The R_Q does not depend on the length and is intrinsic in nature, and R_{mc} is due to the fabrication process and may have a value from zero to several kilo-ohms. The quantitative value of intrinsic conductance (G) can be expressed as [16].

$$G^{-1} = R_Q + R_S \cdot L = \frac{h}{2e^2 N} + \frac{h}{2e^2 N} \cdot \frac{L}{\lambda} \quad (4)$$

where $h/2e^2 \sim 12.9 \text{ k}\Omega$, and λ , L , and N are mfp, the length, and number of conducting channels of each shell, respectively. Additionally, kinetic inductance (L_K) arises due to the kinetic energy of electrons in CNT, whereas magnetic inductance (L_M) is primarily due to the magnetic field associated with CNT. Magnetic inductance is generally neglected as it is of hundredth of the order of kinetic inductance. Kinetic inductance is formulated as [17].

$$L_K = \frac{h}{2e^2 v_F} \quad (5)$$

where

$v_F = 8 \times 10^5 \text{ m/s}$ is known as the Fermi velocity of graphene and CNT.

In RLC model of Fig. 3, quantum capacitance (C_Q) is used to model the density of electronic states in CNT. The electrostatic capacitance (C_E) occurs due to an interaction between ground and CNT, whereas the coupling capacitance occurs between CNTs (C_m) or CNT bundle (C_c). Apart from this, Fig. 2 shows a different kind of capacitance and Fig. 3 shows the equivalent three-line architecture RLC model.

Intershell coupling capacitance is neglected due to its negligible value compared to the quantum capacitance of CNT and can be given as [17–20].

$$C_{Q/\text{channel}} = 2 \times \frac{2e^2}{hv_F} \quad (6)$$

The quantum capacitance per shell can be further defined as

$$C_{Q/\text{shell}} = C_{Q/\text{channel}} \times N_{\text{channel}} \quad (7)$$

Therefore, the quantum capacitance of MWCNT will be the sum of individual shells as

$$C_{\text{MW}}(D_{\text{outer}}, l) = \sum_{d_i=D_{\text{inner}}}^{D_{\text{outer}}} C_{Q/\text{shell}}(D_i, L) \quad (8)$$

Apart from this, the electrostatic capacitance occurs between the CNTs and the ground plane, and coupling capacitance which occurs between neighboring CNTs can be expressed as

$$C_E = \frac{2\pi\varepsilon}{\ln(2y/d)} \quad \text{and} \quad C_C = \frac{\varepsilon\pi l}{\ln\left[\left(\frac{d_c}{2r}\right) + \left(\sqrt{\left(\frac{d_c}{2r}\right)^2 + 1}\right)\right]} \quad (9)$$

where y depicts the height from ground level, d_c denotes the center-to-center distance among any two CNTs, d represents the diameter, l is the length of the CNT interconnect, and r is the mean radius. Furthermore, the coupling capacitance between CNT bundles is given as [17].

$$C_{\text{CM}} = \frac{\pi\varepsilon_0\varepsilon_r}{\ln(S_p/d)} \quad (10)$$

where S_p represents spacing between CNT bundle and d denotes the average diameter of bundled SWCNT and MWCNT [20].

3 Performance Analysis

This section demonstrates the impact of crosstalk-induced delay (signal integrity) for various technology nodes at global interconnect lengths using a multi-line bus architecture.

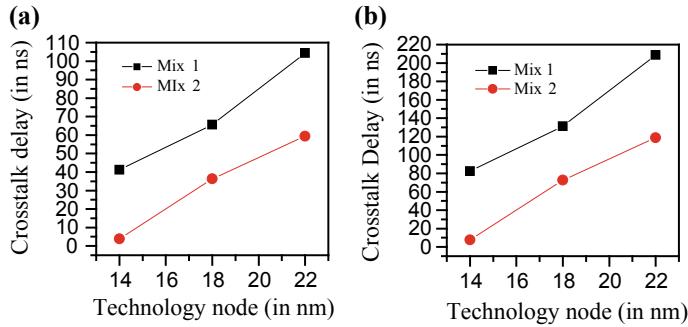


Fig. 4 Crosstalk-induced delay for **a** 100 μm and **b** 200 μm interconnect lengths

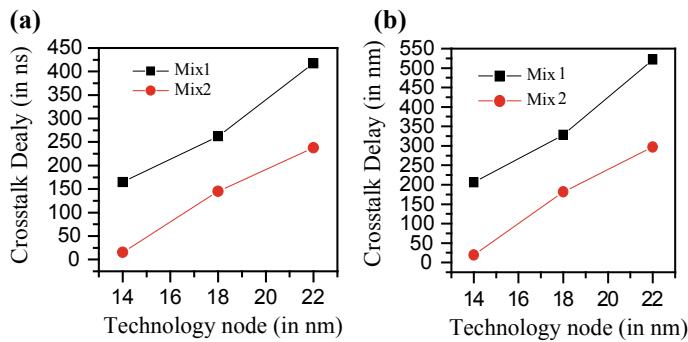


Fig. 5 Crosstalk-induced delay for **a** 400 μm and **b** 500 μm interconnect lengths

3.1 Crosstalk Delay Variation for MCB

In this work, the crosstalk-induced delay for MCB-I and MCB-II structures is obtained at various lengths ranging from 100 to 500 μm and for different technology nodes such as 14, 18, and 22 nm. This research work considers MWCNT with ten shells as compared to DWCNT and SWCNT. The crosstalk delay is analyzed using HSPICE simulator. The increase in delay has been observed for longer interconnects. This is due to the scattering effect that increases with interconnect length. As a result, the quantitative value of interconnect parasitic increases on which delay is proportionally depended. Apart from this, the coupling capacitance has a major effect on crosstalk delay that also depended upon interconnect length and bundle dimensions. Figures 4 and 5 show the crosstalk delay for MCB-I and MCB-II for different technology node, whereas Fig. 6 illustrates the crosstalk delay for varying lengths at particular technology nodes.

The crosstalk delay which becomes worst for 22-nm technology nodes compared to 14 and 18 nm has been observed. This scenario can be understood as the effect of parasitic that primarily depends on the number of CNTs in the bundle. The total

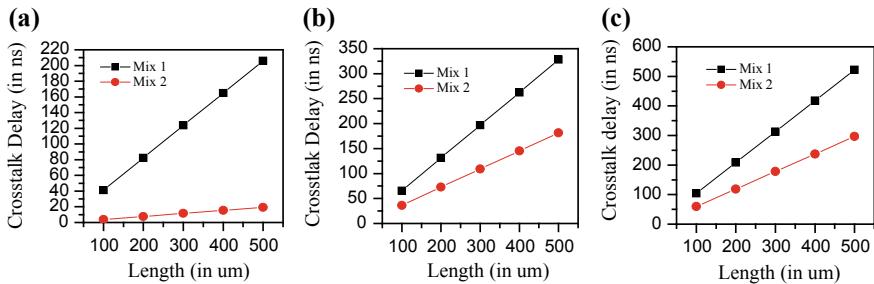


Fig. 6 Crosstalk induced delays for various interconnect lengths at various technology nodes **a** 14 nm, **b** 18 nm, and **c** 22 nm

number of CNTs can be decided from bundle dimension. Higher technology nodes accommodate more CNTs in bundle resulting in a larger value of coupling capacitance. Thus, crosstalk delay becomes more severe.

4 Conclusion

The applicability of MCB has been analyzed at different technology node of 14 to 22 nm for different interconnect lengths of 100 to 500 μm . In this work, the crosstalk-induced delay has been analyzed for MCB and it has been observed that the MCB having MWCNT has significant improvement in crosstalk delay. As technology node increases, MCB-II proved to be more prominent than MCB-I as former is providing less crosstalk as compared to latter. The comparison between crosstalk delay of MCB-I and MCB-II shows an improvement of 6.8% approx.

References

1. Kaushik, B.K., Majumder, M.K., Kumar, V.R.: Carbon Nanotube based 3-D interconnects—a reality or a distant dream. *IEEE Circ. Syst. Mag.* **14**(4), 16–35 (2014)
2. Li, H.: Carbon nanomaterials for next-generation interconnects and passives: physics, status, and prospects. *IEEE Trans. Electr. Devices* **56**(9), 1799–1821 (2009)
3. Xu, Y., Srivastava, A.: A model for carbon nanotube interconnects. *Int. J. Circ. Theory Appl.* **38**(6), 559–575 (2010)
4. Kumar, V.R., Majumder, M.K., Kaushik, B.K.: Graphene based on-chip interconnects and TSVs—prospects and challenges. *IEEE Nanotechnol. Mag.* **8**(4), 14–20 (2014)
5. International Technology Roadmap for Semiconductors: (2012), [Online]. Available: <http://public.itrs.net>
6. Im, S., Srivastava, N., Banerjee, K., Goodson, K.E.: Scaling analysis of multilevel interconnect temperatures for high performance ICs. *IEEE Trans. Electr. Devices* **52**(12), 2710–2719 (2005)
7. Yokogawa, S., et al.: Tradeoff characteristics between resistivity and reliability for scaled-down cu-based interconnects. *IEEE Trans. Electr. Devices* **55**(1), 350–357 (2008)

8. Dadgour, H., Cassell, A.M., Banerjee, K.: Scaling and variability analysis of CNT-based NEMS devices and circuits with implications for process design. In IEDM Tech. Dig. pp. 529–532 (2008)
9. Li, H., Xu, C., Srivastava, N., Banerjee, K.: Carbon nanomaterials for next-generation interconnects and passives: physics, status and prospects. *IEEE Trans. Electr. Devices* **56**(9), 1799–1821 (2009)
10. Subramaniam, C.H. et al.: One hundred fold increase in current carrying capacity in a carbon nanotube–copper composite. *Nature Commun.* **2**, 2202 (20013)
11. Han, Zh, Fina, A.: Thermal Conductivity of carbon nanotubes and their polymer nanocomposites: a review. *Prog. Polym. Sci.* **36**(7), 914–944 (2011)
12. Berber, S., Kwon, Y.-K., Tomanek, D.: Unusually high thermal conductivity of carbon nanotubes. *Phys. Rev. Lett.* **84**(20), 4613–4616 (2000)
13. Avouris, P., Chen, Z., Perebeions, V.: Carbon-based electronics. *Nat. Nanotechnol.* **2**(10), 605–613 (2007)
14. Subash, S., Kolar, J., Masud, H.: A new spatially rearranged bundle of mixed carbon nanotubes as VLSI interconnect. *IEEE Trans. Nanotechnol.* **12**(1), 3–12 (2011)
15. Naeemi, A., Meindl, J.D.: Compact physical models for multiwall carbon-nanotube interconnects. *IEEE Electr. Devices Lett.* **27**(5), 338–340 (2006)
16. Li, H., Yin, W.Y., Banerjee, K., Mao, J.F.: Circuit modeling and performance analysis of multi-walled carbon nanotube interconnects. *IEEE Trans. Electr. Devices* **55**(6), 1328–1337 (2008)
17. Burke, P.J.: Luttinger liquid theory as a model of the gigahertz electrical properties of carbon nanotubes. *IEEE Trans. Nanotechnol.* **1**(3), 129–144 (2002)
18. Majumder, M.K., Pandya, N.D., Kaushik, B.K., Manhas, S.K.: Dynamic crosstalk effect in mixed CNT bundle interconnects. *IET Electr. Lett.* **48**(7), 384–385 (2012)
19. Kaushik, B.K., Sankar, S., Agarwal, R.P., Joshi, R.C.: An analytic approach to dynamic crosstalk in coupled interconnects. *Microelectronic. J.* **41**(2/3), 85–92 (2010)
20. Majumder, M.K., Kaushik, B.K., Manhas, S.K.: Analysis of delay and dynamic crosstalk in bundled carbon nanotube interconnects. *IEEE Trans. Electromagn. Compat.* **56**(6), 1666–1673 (2014)

Assignment Tracking on Android Platform



Dhriti Chakraborty and Matangini Chattopadhyay

Abstract Student assessment in terms of ‘assignment tracking’ is essentially needed in order to understand students’ cognition level in a learning process. Students are asked to write programs to solve problems, and then, those are evaluated. The traditional offline method of this formative assessment by pen-and-paper mode is time-consuming and tedious. There are many online assessment methodologies too. Android OS-based smartphones becoming so popular are being effectively used in the domain of education. Online formative assessment process for Java programming on android platform is proposed in this paper. The ‘Assignment Tracker’ application is developed and designed using official IDE, Android Studio 3.1. Android OS supports the Dalvik bytecode instead of traditional JVM. So, a client–server architecture is proposed. The students on client side can execute Java programming on their android mobiles and request the server for compiling and execution. A third-party Web server is utilized to build up the connection between client and server. On implementation, the overall system proves fruitful and effective as an advanced formative assessment technique.

Keywords Assessment · Android OS · Java programing · IDE · JVM · Client–server architecture · Spring framework · Third-party web server

1 Introduction

In the last few years, the android OS-based mobile phones have captured the major part of the smartphone market because it is Linux-based and open source. Most of the applications running on android OS are written in Java. Java is general-purpose computer programming language. It is object-oriented and portable. Java has created

D. Chakraborty (✉) · M. Chattopadhyay
School of Education Technology, Jadavpur University, Kolkata, India
e-mail: dhrimon22@gmail.com

M. Chattopadhyay
e-mail: matanginic@gmail.com

a new paradigm in software development, and learning this language is an important module in software development.

In a Java program learning course, it is necessary to assess the student's cognition level by asking them to solve programming problems. The traditional method of programming assessment is offline based on pen-and-paper mode. As the number of students and the working load have increased, the offline assessment proves to be time-consuming, error-generating, and tedious [3]. There are many online assessment methodologies for assessing programming skills [5–9]. To develop a Java learning module in android platform, a Java programming assessment method is required for android OS. The students have to execute Java programming in response to given problems on their smartphones. A compiler is needed to check the written coding and get corresponding O/P. The major problem to be faced in this aspect is that android supports the Dalvik bytecode, instead of traditional JVM.

With this view, client-server-architecture-based online formative assessment system for Java programming on android platform is proposed in this paper. The 'Assignment Tracker' app is developed for android where students on client side can execute their assignments on Java programming on their smartphones. The app is provided with the facility to compile the written coding. The request of compiling is sent to the server; the server executes the code and sends the O/P accordingly. Another major concern in this aspect is the connection between the client and the server because of Windows Firewall. A third-party Web server is utilized to build up the connection.

Rest of the paper is organized as follows: Sect. 2 includes the existing online assessment techniques regarding programming languages. Section 3 outlines the proposed approach. Section 4 describes the implementation of the proposed system. In Sect. 5, we conclude after discussing the future scope.

2 Related Work

Astin [1] introduced the term student involvement, which is defined as 'the amount of physical and psychological energy that the student devotes to the academic experience.' According to Race [2], 'The most important thing we do for our students is to assess their work.'

Conventional methods of assessing programs manually are tedious. The increase in the number of students on courses produces enormous load on the evaluator. Again, sharing this burden between two or more tutors arises the question of inconsistency and incompatibility. Even individual tutor may be inconsistent while evaluating the same assignments of different students [3].

Many automated evaluation tools for assessing programming skills use a set of test (input) data and a set of benchmark results [4]. These tools try to execute programs on test input data and compare the results with the benchmark results.

It is also possible that the students will submit their assignments electronically. The evaluator may use a batch or script file to compile the submissions and execute the program. The generated output is automatically checked against the benchmark

data. Examples of such script-based marking systems are Ceilidh [5] and CAAPE [6].

Thorburn and Rowe [7] developed a system called Program Assessment using Specified Solutions (PASS). This system is used to evaluate programs written in C.

Another evaluation system JAssess has been proposed by Yosuf et al. [8]. This semiautomatic system is integrated with Moodle and is used to evaluate Java programs submitted in Moodle.

RoboProf [9] proposes a smart evaluator of programming skills. The programming assignments are presented to the assessee in the order of increasing difficulty. Thus, a student can self-evaluate their progress.

Another Web-based assessment system is BOSS [10]. In this online submission and grading are allowed. Even online feedback is also possible. BOSS is used to mark programs in C.

The above mentioned is a brief discussion on a few important developments in the aspect of student assessment in programming courses. With these previous studies, the motivation of designing a system for online assessment of Java programming on android platform puts us forth to the present study.

3 Proposed System

The proposed assignment tracking system is based on client–server architecture. The ‘Assignment Tracker’ application is developed and implemented for android-based smartphones on client side. The application is designed using the official integrated development environment (IDE), Android Studio 3.01. The basic features of the application are shown in Fig. 1.

The ‘Start’ button provided at the middle of the screen allows the user to start writing the assignment code. The question comes on the title bar, and a timer starts simultaneously on clicking the ‘Start’ button. Users are provided with ‘Save,’ ‘Delete,’ and ‘Compile’ buttons to avail the features accordingly. The ‘Delete’ and ‘Compile’ options are active only for the saved assignments.

Figure 2 shows the Level 0 DFD of the application. When the user launches the application and clicks on the ‘Create’ button, he/she is provided with a working space along with Start button. When user clicks on the ‘Start’ button, the assignment evaluation context view opens allowing him/her to write the Java code in response to the question appeared in the title bar along with a timer. User clicks on the ‘Save’ button after he/she has completed the assignment. The document is saved showing the ‘Home’ page with list of saved documents. When user clicks on the ‘Compile’ button, opening the saved assignment, the written Java code is sent to the server and compiled. The application shows the O/P accordingly.

The entire system architecture is shown in Fig. 3. When the client or user requests for compiling the written Java code as per given assignment, the written code is encoded using base 64 algorithm. Then, the encoded code is sent to the server. It is to

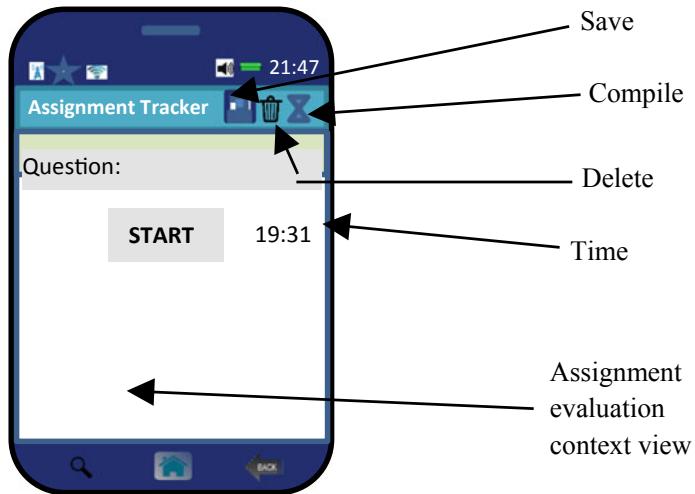


Fig. 1 Basic features of assignment tracker app

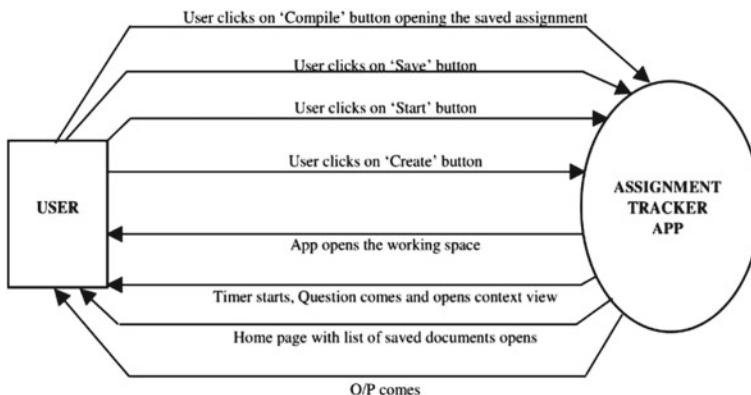


Fig. 2 DFD Level 0

be noted that in the entire procedure the clients are connected to the server through cellular networks like 2G, 3G, or 4G.

The client side in this aspect cannot be directly connected to the server due to Windows Firewall. To build up the connection, a third-party Web server ‘ngrok’ is utilized. It exposes local servers behind firewalls to the public Internet over secure tunnels. It receives the encoded code from client side and sends it to the server for compilation.

The server is written in Java and developed using Spring Framework. The server receives the encoded code from the client through third-party ngrok; then, it decodes the code using base 64 algorithm. The Java coding is compiled on the server using

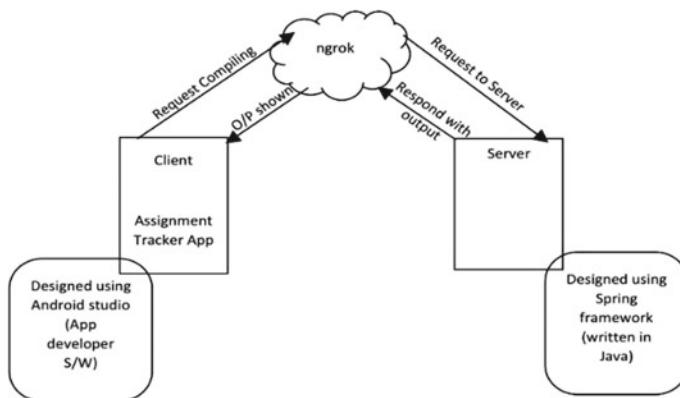


Fig. 3 Schematic representation of system architecture

string compilation procedure. The server processes the code, produces the O/P as string, and then sends it to the user through ‘ngrok’.

4 Implementation and Results

The ‘Assignment Tracker’ app is designed and developed for mobile devices which are based on android OS. The proposed system is developed and implemented using a laptop with the following configurations.

- 4 GB RAM
- 1 TB hard disk
- Intel Core i3 processor
- Windows 8.1 Pro operating system.

The software requirements for developing the app are

- JDK version 8.0.1710.11
- Android Studio version 3.0.1 with Android SDK version 24
- STS version 3.9.4.

A virtual device (Google Nexus 5) is created as an emulator with the help of Android Virtual Device (AVD) manager. To send the code from client side to the server side, Internet connection is required. For this, we have to access the Internet.

<uses-permission android:name="android.permission.INTERNET"/>

Different activities are added to the student tracker app. Their names and labels are added under the application tag using activity tag as shown below.

</application>

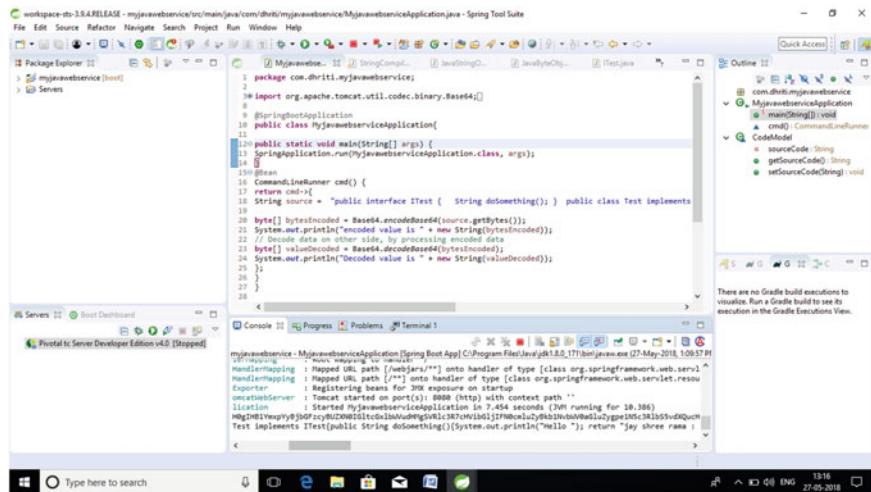


Fig. 4 Screenshot showing the server has started

```
<activity android:name="MainActivity"
    android:label="Assignment Tracker APP"/>
<activity android:name="NoteActivity"
    android:label="Assignment Tracking page"/>
<activity android:name="CompileActivity"
    android:label="OUTPUT"/>
<activity android:name=".
Activity_Help" android:label="Help"/>
</application>
```

Similarly, more activities are added and given relevant names and labels. The layouts and icons are designed in such a way that gives a smooth experience.

First, the server is run using the Spring boot app from ‘myjavawebservice’. The server is started, and now, it has to be connected with the application. The ‘ngrok’ gets the URL of the server and gives it to the application. The screenshot showing the start of the server is shown in Fig. 4.

The ‘Assignment Tracker’ app is installed on the smartphone on the client side. Figure 5 shows the home page of the application. User is provided with ‘Create’ button on the right-hand corner of the home page. On clicking the ‘Create’ button, the user is navigated to the assignment evaluation page. On the assignment evaluation page, the student has to click the ‘Start’ button to begin the assignment as shown in Fig. 6. A question comes on the title bar, and the timer starts simultaneously. The user has to write the Java code corresponding to the question within the doSomething() command.

On completion of the assignment, the user can save the same by clicking the ‘Save’ button and the user is directed to the home page showing the saved document

Fig. 5 Screenshot showing the ‘home page’

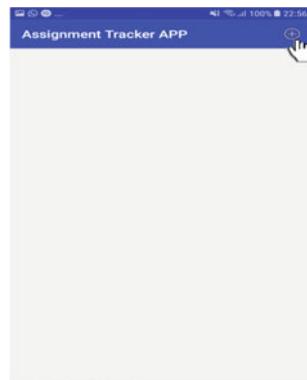
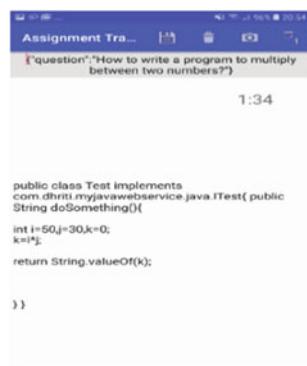


Fig. 6 Screenshot showing the ‘assignment evaluation page’



Fig. 7 Screenshot showing the option to save



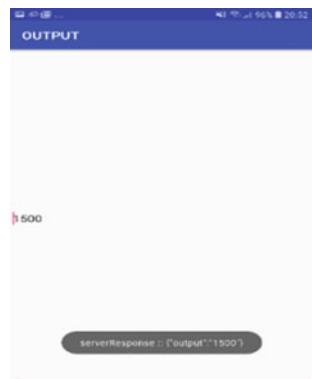
```
public class Test implements  
com.dhruvi.myjavawebserice.java.ITest{ public  
String doSomething(){  
int i=50,j=30,k=0;  
k=i*j;  
return String.valueOf(k);  
}}
```

in the list of saved documents. Otherwise, the assignment will be automatically saved when the time is up. Figure 7 shows that the assignment is saved, and Fig. 8 shows the list of saved documents.

Fig. 8 Screenshot showing the saved document with the list view



Fig. 9 Screenshot showing O/P from the server

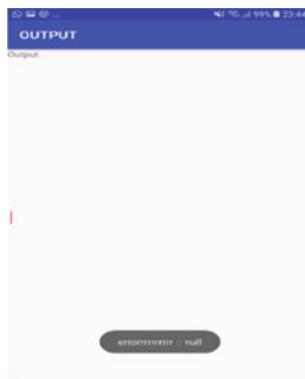


The user has to open the saved assignment in order to compile. When the user clicks the ‘Compile’ button, the request is sent to the server and the code is executed. The response from the server comes to the app with corresponding O/P as shown in Fig. 9. A null message comes as the O/P in case of error in the code which is shown in Fig. 10. The user is also provided with the option for deleting a document.

5 Conclusion and Future Scopes

‘Assignment Tracker’ app so designed may be fruitfully used as an advanced formative assessment technique in Java programming learning module. Student assessment with time limit can be done effectively by the proposed system. It is mandatory that both the server and the client are online during the execution of assessment procedure. It is observatory to note that the assignments once saved, intentionally or completion before time or forcefully due to finish of the scheduled time, cannot be modified further but compiled and deleted only. The student has to compile in order

Fig. 10 Screenshot showing the null O/P from the server in case of error



to submit their assignments. So, the tracking procedure is effective and advantageous to many of the traditional techniques.

The present study deals with the assessment of Java programming learning. This app may be developed in future with the provision of assessments of other programming languages. The use of a front camera during the execution of assessment procedure can also be developed so that the instructor can detect the face movement of a student from the server side. Future works on this front camera will provide an excellent assessment technique where the server side will have a choice to warn a student or take some step if a notorious movement is detected during assessment.

References

1. Astin, A.W.: Student involvement: a developmental theory for higher education. *J. Coll. Student Dev.* **40**(5), 518–529 (1999)
2. Race, P.: Designing Assessment to Improve Physical Sciences Learning: A Physical Sciences Practical Guide. Higher Education Academy (2009)
3. Edmunds, G.: Experiences using CAAPE: computer assisted assessment of programming exercises. *Comput. Educ.* **15**(1–3), 45–48 (1990)
4. Ala-Mutka, K.: A survey of automated assessment approaches for programming assignments. *Comput. Sci. Edu.* **15**(2), 83–102 (2005)
5. Benford, S., Burke, E., Foxley, E., Gutteridge, N., Zin, A.M.: Experience using the Ceilidh system. In: Proceedings of the 1st All-Ireland Conference on the Teaching of Computing, pp. 32–35 (1993)
6. Edmunds, G.: Experiences using CAAPE: Computer assisted assessment of programming exercises. *Comput. Educ.* **15**(1–3), 45–48 (1990)
7. Thorburn, G., Rowe, G.: PASS: an automated system for program assessment. *Comput. Educ.* **29**, 195–206 (1997)
8. Yusof, N., Zin, N.A.M., Adnan, N.S.: Java programming assessment tool for assignment module in moodle e-learning system. *Procedia Soc. Behav. Sci.* 767–773 (2012)
9. Daly, C., Waldron, J.: Assessing the assessment of programming ability. In: Proceedings 35th Technical Symposium on Computer Science Education, pp. 210–213 (2004)
10. Joy, M., Griffiths, N., Boyatt, R.: The boss online submission and assessment system. *J. Educ. Resour. Comput.* **5**(3), 1–28 (2005)

Empowering Textile Crafts Through the Internet of Things Technology



Deepshikha and Pradeep Yammiyavar

Abstract Textile traditions of India although rich and varied have been on a gradual decline over the past decades. With technology becoming a part of human beings through lifestyle products and services, it is essential to intervene a segment of traditional crafts with smart technology for increased competence in the global market. This paper discusses in brief the state-of-the-art research being conducted globally on e-textiles in biomonitoring, military, home interior and clothing sectors and their implications in design intervention of traditional textile crafts of India embedded with functions combined with cultural aesthetics. It further presents a digital retail concept for selling handloom and handcrafted products with fair share given to the weavers for digital sensitization in the craft sector. The third segment of the paper discusses creation of a digital repository for handcrafted products which can be accessed through an external device explained via an information flow diagram. The paper thus focuses on need for technological intervention in the handcrafted sector through applications of Internet of Things and other smart technologies available at present through conceptual examples and their implications.

Keywords Textile crafts · Digital retail · Digital repository

1 Introduction

The textile industry in India is one of the largest sources of employment generation with 45 million people involved in allied activities with a net worth of 150 billion USD [1]. It contributes 4% to the GDP and 14% to industrial production in India [1]. The handloom sector employs 4.3 million people across the country being the second largest occupational sector after agriculture for the rural India [2]. Although

Deepshikha (✉) · P. Yammiyavar

Department of Design, Indian Institute of Technology, Guwahati, Assam, India

e-mail: deepshikhaphd@iitg.ac.in

P. Yammiyavar

e-mail: pradeep@iitg.ac.in

the export from handloom sector was US\$ 357.53 in 2016–17 [2], handloom as a sector has been declining sharply [3] mainly due to the following factors—low productivity compared to the mill sector, limited scope of technological upgradation, skilled labours migrating to other occupations, seasonal fluctuations, slow growth and production, low wages, work involves high physical activities and others. Over the last decade, several government and NGOs have provided support for development, training programmes, advertising and facilitation of sale through local and export retailers [3] which has resulted in better representation of the traditionally crafted products on retail and other platforms. Similar trends are seen in other handcrafted products apart from handloom where traditionally made crafts are being practiced in fewer pockets than before due to rapid industrialization and globalization over the past decades.

With technology becoming a part of daily lifestyle [4, 5], textiles have become an interesting and useful medium for embedding, collecting and transmitting information in several domains such as health, sports, military, security, interiors, clothing and automobile. The present paper discusses few state-of-the-art examples of e-textiles and possibilities of merging traditionally crafted textiles with technology for smart applications which could lead to preservation of hierarchical crafts and bring them at par with technologically advanced e-textiles of present. The paper elaborates a retail platform for crafts for a prospective digital retail revolution at grass root level and creation of a digital information repository of crafts connected through the Internet of Things (IoT) technology.

2 State-of-the-Art Example of E-Textiles

WEALTHY (Fig. 1) is a health monitoring system for cardiovascular diseases comprising sensors and electrodes with advanced signal processing capabilities for real-time assessment [6]. Figure 2 depicts military uniform with e-textiles by BAE Boardsword [7] comprising woven smart threads, embedded computer for power and data, embedded battery storage with flexible charger, fabric-mounted connectors in customer-defined locations, head-mounted displays and other wearables connected with the uniform for wireless transmission of data to the control station.

Cutecircuit (Fig. 3) has made a series of led-based textiles for occasional wear, fashion couture and performances. For a British airline carrier, Easyjet (Fig. 4), Cutecircuit has designed LED based uniform [8] that provides the cabin crew with extra lighting, a lapel that displays flight information and built-in microphone for easier communication between the crew members. Textales [9] has embedded fairytale application on the bedsheet for children's edutainment (Fig. 5). Textales has been developed as part of series of smart textile products by the Creative Industry Scientific Programme, a Dutch government-funded initiative that collaborated with industrial and scientific partners. Oerlikon in its press release [10] for Techtextile 2015, Frankfurt, mentions that textiles for automobiles in near future may employ conductive

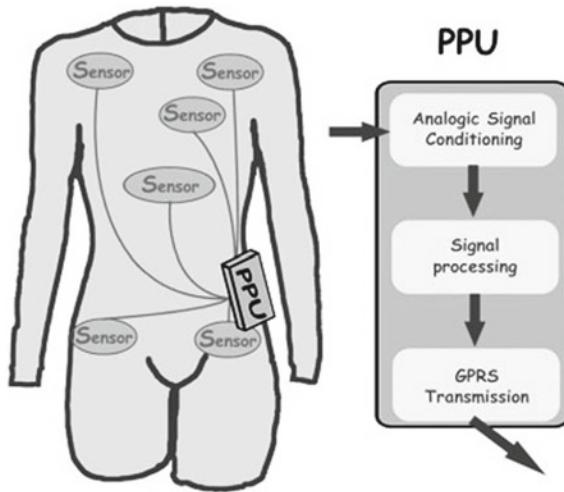


Fig. 1 WEALTHY [6]

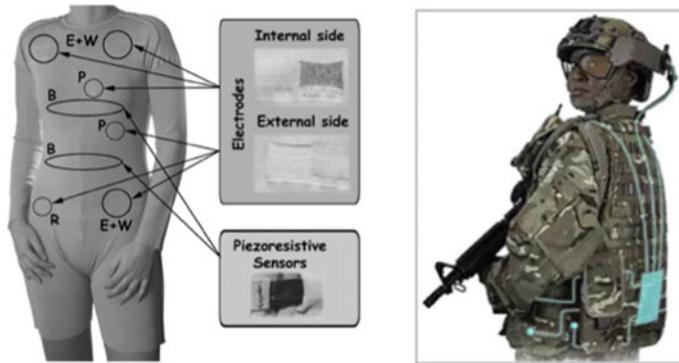


Fig. 2 Soldier wearing smart uniform [7]

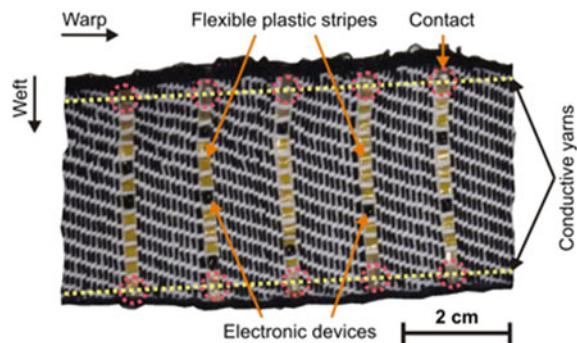
fabrics for cooling and heating, responsive textile panels that can convert gestures to gear shifting, biomonitoring of driver or passengers and other applications.

2.1 Implications for Designing Handcrafted E-Textiles of India

Wearable electronics such as Lilypad Arduino, conductive threads, sewable micro LEDs, conductive inks, sensors and temperature sensitive inks can be integrated with traditional designs seamlessly to create technologically competent products. Such

Fig. 3 Cutecircuit**Fig. 4** Easyjet uniform

products need not be mass produced on looms or by artisans, but may be designed and developed as a specific line of products by design studios, brands, NGOs, etc. or initiated by the government as upgradation programmes. For example, block printed fabrics may be developed with thermochromic pigments that could change patterns with change in temperature. Sensor strips could be woven with the handloom yarns as warp or weft for conductive handloom woven textiles. As depicted in Fig. 7—a handloom woven fabric has been developed by researchers at University of Laval for biomonitoring and military applications [12]. In Fig. 6, researchers have woven flexible plastic strips to embed LEDs and photodiodes for near-infrared spectroscopy [11]. Lilypad Arduino [13, 14] developed by MIT researchers is a microcontroller that can be easily embroidered or integrated with fabrics, connected to sensors,

Fig. 5 Textales CRISP**Fig. 6** Woven smart fabric [11]

actuators and Arduino coded for multiple functions (Fig. 8). Based on state-of-the-art research review, traditionally crafted products could be intervened and embedded with smart technology for augmented functions to aid users in daily life and in several domains, such as health, geriatric, to meet the needs of the specially abled, military, home interiors and automobile in a more Indian context. This would not only preserve hierarchical traditions but also provide a cutting edge to traditionally crafted products and educate groups of artisans to create technologically competent designs with value addition.

3 Role that Digital Retail Can Play in Empowering Craftsmen

Traditional business methods surrounding crafts sector involve a number of intermediaries between the maker of the product and buyer of the product. With margins becoming less and with sale of crafts having to compete with machine manufactured

Fig. 7 Woven sensors,
university of laval [12]

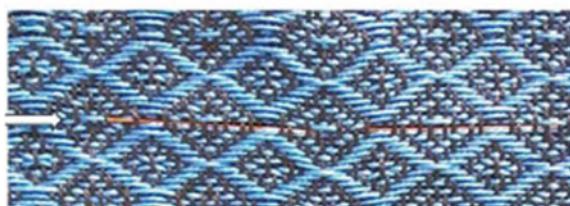
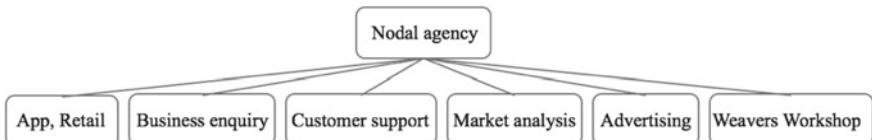


Fig. 8 Lilypad Arduino
embroidered on fabric [13,
14]



goods, such large network of sales and distribution becomes unviable. There have also been reported instances of the artisans being denied their fair share of accruals by the middleman in business. This part of the paper discusses development of a conceptual digital platform specifically designed for small craft clusters.

In present scenario, weavers usually sell from home, contact local buyers in nearby town and cities. They participate in handloom fairs, etc. but are usually expensive for them. The price in showrooms or retail outlets is much higher than the purchased rate at which the weaver supplies, which diminishes the cost payable to individual weaver. Weavers themselves are not aware or making use of retail concepts such as branding, advertising and online selling. They would hence require a moderating unit that would act as a control agency to train them in concepts of digital retail at a grass root level and connect them to the customers through technological upgradation (Fig. 9). The control agency could be a government body or an NGO or a self-

**Fig. 9** Role of control unit

STRENGTHS	WEAKNESS	OPPORTUNITIES	THREATS
Traditional heritage	Slow production	Youth marketing	'Middlemen'
Wide and varied	Seasonal fluctuations	Global marketing	Power-loom
Variety of products	Expensive	IOT services	Wash care sensitive
Style statement	Poor sensitization	Digitisation	Slow sector
Unique identity	Limited technological intervention	Design studios	Competition

Fig. 10 SWOT Analysis**Fig. 11** Conceptual application graphical user interface (GUI)

help group initiative. This body could take care of product specifications, packaging, business enquiries, customer care, logistics, festive offers, discounts, type of products to be showcased, inventory, etc. Figure 11 depicts graphical user interfaces (GUI) of a conceptual application for handicraft sector controlled by the nodal unit. Few images of Fig. (11) GUIs have been used as reference only from Fabindia [17].

A conceptual mobile application has been designed to connect the buyer directly to the seller, thereby maximizing profit to the maker of the craft goods. The system requires a controlling agency to be situated in the nodal craft cluster that could help in sensitization and providing minimal technical literacy to the weavers for basic understanding of the retail platform, business model keeping pace with retail trends.

and the benefits that they could receive. This platform defines the geographical area that could access the nodal unit to source the product for simpler logistics. Bulk buyers are usually located in nearby towns or villages that could benefit apart from other users for whom doorstep delivery could be a more feasible option than travelling to the place of manufacture. Secondly, since it ensures that middlemen will not be involved, healthy margins could directly be received by the weavers. Thirdly, since the key players of the online retail market sell to a category of urban and semi-urban users, there is a niche to serve to the semi-rural, rural–urban and also urban users. The set-up follows both business-to-business and business-to-customer approach for increased benefit to the weavers. Through the involvement of the nodal cluster office, the business model and the mobile application, the researchers aim to make the rural weavers independent, responsible sellers, introduce to them a modern platform of commerce and gradually revolutionize selling in the handicraft sector. Using design methods in interaction design and graphic user interfaces for the platform are depicted in detail to trace the working of the system. Figure 10 presents a SWOT analysis of the handicraft sector.

4 Creation of a Digital Repository for Handloom and Handicrafts

A traditional textile or craft embedded with information can be accessed by any other person (Figs. 12 and 13) to know about its craftsmanship, traditions, history, place of purchase, how to purchase, nearest place to purchase, easiest/most affordable place to purchase, browse for similar products, know the users style statement, profile if allowed to access or to connect, etc. The data repository needs to be created by an authentic government approved body after thorough historical and state-of-the-art analysis for larger segment of users to know about crafts that are worn by others or placed at a certain place, instantaneously. This will help in propagation of hierarchical knowledge and also promote the products likability and willingness to purchase through retail platforms. In an extended scenario, facilitated by the IoT technology, one could purchase a product just seen in a place worn/owned by another person promptly. Digital repository could contain information as texts, images, videos, augmented reality and virtual reality.

5 Discussion

A technology acceptance model [15]-based study was conducted with 30 young respondents studying at graduate and postgraduate level to understand—(i) if they would find above-mentioned digital platforms useful for gathering information; (ii) if learning to operate these platforms would be easy for them; (iii) if their interaction

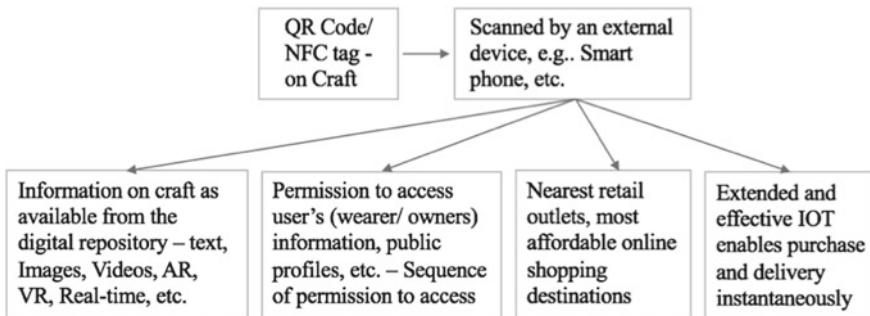


Fig. 12 Information flow for digital repository for handcrafted products



Fig. 13 Scanning textile craft for accessing information through a digital repository

with the conceptual interfaces was clear and understandable; (iv) if it would be easy for them to become skilful at using these platforms; (v) if interacting with conceptual interfaces was easy to use; and (vi) if they are willing to use such platforms. Respondents were to rate the above factors on a Likert scale of 1–7 (1 = low, 7 = high) after interacting with the conceptual GUI screens on an android device in a brief questionnaire administered to them. Respondents revealed that the perceived ease of use pertaining to the digital platforms was high and that they found it to be useful in propagating craft traditions. Figure 14 summarizes mean, standard deviation and mode values of the response.

Researchers [16] have mentioned that as textile production shifted from hand-crafted to industrialized production, apart from power mill sector, handloom, customized tailoring, designers and studios, in near future smart technology will also contribute significantly as wearables and textiles in different domains (Fig. 15). Textiles will become responsive and dynamic sources of information gathering, processing and transmitting through a Web of IoT-controlled segments. As researchers abroad have been exploring smart textile at a faster pace over the last decade, Indian

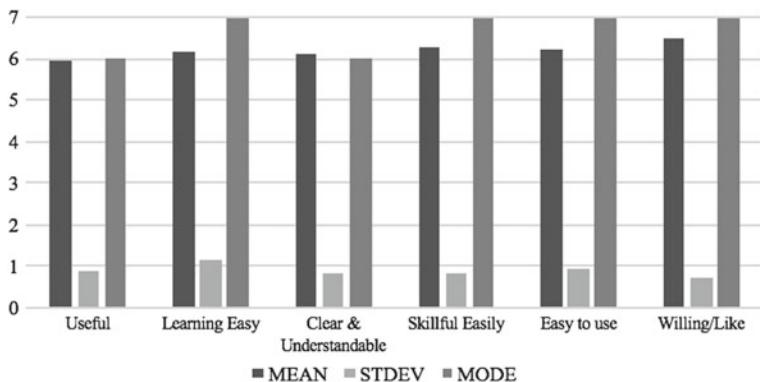


Fig. 14 Perceived ease of use of digital platforms among respondents



Fig. 15 Timeline for textile trends in India [16]

researchers need to develop domain-specific augmented textiles combined with function and aesthetics. The rich knowledge base of traditional crafting may benefit by integration of traditional methods with smart materials for preservation and propagation of traditional heritage. The present paper presents a brief discussion of contemporary e-textile advancements, possibilities of merging traditional crafts with smart technology, conceptual digital retail platform and conceptual digital repository of crafts to empower the traditional textile and craft sector. The paper encourages researchers, designers, academicians and policy makers to probe for identifying and implementing feasible opportunities.

References

1. India Brand Equity Foundation—Textile Industry in India, <https://www.ibef.org/industry/textiles.aspx>. Last accessed 2018/06/27
2. India Brand Equity Foundation—Handloom sector in India, <https://www.ibef.org/exports/handloom-industry-india.aspx>. Last accessed 2018/06/27
3. Note on handloom sector—Ministry of Textiles, Government of India, <http://handlooms.nic.in/writereaddata/2486.pdf>. Last accessed 2018/06/27

4. Schneiderman, D., Winton, A.G.: Textile Technology and Design: From Interior Space to Outer Space. Bloomsbury Academic, London, UK (2016)
5. Quinn, B.: Textile Futures: Fashion, Design and Technology. Berg, Oxford, NY, USA (2010)
6. Paradiso, R., Loriga, G., Taccini, N.: A wearable health care system based on knitted integrated sensors. *IEEE Trans. Inf. Technol. Biomedicine.* **9**(3) (2005)
7. Wearables at war- How smart textiles are lightening the load for soldiers. <https://www.cnet.com/news/wearables-at-war-how-smart-textiles-are-lightening-the-load-for-soldiers/>. Last accessed 2018/06/27
8. Easyjet: <https://www.wearable.com/wearable-tech/easyjet-cutecircuit-for-smart-uniforms-1934>. Last accessed 2018/06/27
9. CRISP Textales: <http://www.crisprepository.nl/project/smart-textile-services/prototype/textales>. Last accessed 2018/06/27
10. Oerlikon Barmag and Oerlikon Neumag- Techtextil (2015). https://www.oerlikon.com/ecomaXL/files/oerlikon_PR-OMF_The_car_of_the_future_will_be_wearing_more_textiles_en.pdf&download=1. Last accessed 2018/06/27
11. Christoph, Z., Nasser, N., Lars, B., Münzenrieder, N., Kinkeldei, T., Petti, L., Kleiser, S., Salvatore, G.A., Wolf, M., Tröster, G.: Textile integrated sensors and actuators for near-infrared spectroscopy. *Opt. Express* **21**, 3213–3224 (2013)
12. Smart woven textile—University of Laval, <https://www.pinterest.co.uk/pin/311381761711368300/>. Last accessed 2018/06/27
13. Buechley, L., Eisenberg, M., Catchen, J., Crockett, A.: The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08), pp. 423–432. ACM, New York, NY, USA (2008)
14. Becky Stern Lilypad Embroidery. <https://beckystern.com/2008/04/21/lilypad-embroidery/>. Last accessed 2018/06/27
15. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, **13**(3), 319–340. Management Information Systems Research Center, University of Minnesota, USA (1989)
16. Deepshikha, Yammiyavar, P., Nath, N.: Smart textile trends and their implications in digitizing craft traditions. In: Proceedings of International Conference on Recent Trends and Sustainability in Crafts and Design- A treatise on recent trends and sustainability in crafts and design (IICD 2017), pp. 140–147. Excel Publishers, Jaipur, India (2017)
17. Fabindia. <https://www.fabindia.com/>. Last accessed 2018/06/27

RSA Using Montgomery Powering Ladder on Dual Core



Pawan Singh, Ashwani Kumar and Mohit Kumar

Abstract RSA is widely used public key cryptographic algorithm. Security of RSA depends upon the factorization of the large integers. Key sizes must be large for highly secured RSA. But the computation time of RSA increases for larger key sizes. Sequential RSA encryption for 1024 bits takes large execution time. Montgomery powering ladder prevents simple power attack and can be parallelized. In this paper, parallel Montgomery powering adder is implemented on dual-core machine using the multiprocessing module of the python. The performance of the parallel Montgomery powering ladder is analyzed for various key sizes. The parallel Montgomery powering ladder shows 1.63 speedup over its sequential version.

Keywords RSA · Montgomery powering ladder · Dual core · High-performance RSA · RSA encryption · RSA decryption · FIOS · Multiprocessing

1 Introduction

Internet is a worldwide used, efficient, user-friendly, and economical medium of communication. Nowadays, it is used in every field like educational institutions, private organizations, military, and financial institutions. These organizations may transfer critical data which requires confidentiality. There is possibility that message can be accessed during the transmission. Cryptographic algorithms prevent unauthorized access by encrypting the data.

P. Singh

Department of CSE, Pranveer Singh Institute of Technology, Bhauti,
Kanpur 209305, India

A. Kumar (✉)

Department of CSE, Vardhaman College of Engineering, Hyderabad,
Telangana, India
e-mail: ashwani.kumarcse@vardhaman.org

M. Kumar

Department of CSE, Cambridge Institute of Technology, Ranchi, India

Public key cryptographic algorithms are important because it provides the opportunity to communicate even if the parties do not have the method to agree on common secret key. The two different keys are used by these algorithms are public and private keys, for the purpose of encryption and decryption, respectively; the security of these algorithms depends upon the keeping the private key secret.

RSA is widely used public key cryptographic algorithm [1]. Security of RSA depends upon the factorization of the modulus. Encryption and decryption in RSA are computationally intensive operations because they make use of large number's modular exponentiation. To ensure the security of the RSA, the key size must be 1024 bits or greater. As the size of key increases, the computational cost also grows.

In the current scenario, general-purpose processors have more than one core that performs computations simultaneously [2, 3]. This parallel hardware causes the software designer to reconsider the design of software for efficient utilizations of the resources [2, 3]. Parallel programming on different cores is major challenge in the transition from single-core to multicore systems. Programming environments like multiprocessing module in Python provide the parallel programming models to abstract the underlying multicore architecture [4, 5].

In this paper, the performance of RSA is enhanced by utilizing the core of dual-core processor. Section 2 discusses the computations of RSA cryptosystem. Section 3 presents the proposed approach for implementation of RSA on dual core. Section 4 gives the implementation results and comparison. Section 5 gives the conclusion and future scope.

2 RSA Algorithm

RSA is widely used public key cryptosystem which was introduced by Leonard Adleman, Adi Shamir, and Ron Rivest in 1977 at MIT [1]. It is one of the most important algorithms used for encryption and authentication for communications on Internet. It was the first suitable algorithm for both digital signature and data encryption. It is widely used in protocols supporting the e-commerce.

The RSA cryptosystem has three computations:

1. Generation of key, i.e., key generation
2. Algorithm for encryption
3. Algorithm for decryption.

2.1 Key Generation

The generation of key is a process in RSA algorithm of multistep process which is given below as follows [6]:

1. Choose two prime numbers p and q which are very large.

2. Compute the modulus $n = p * q$
3. Compute the $\phi(n) = (p - 1) * (q - 1)$ where $\phi(n)$ is Euler's totient.
4. Then an integer e is chosen, such that $1 < e < \phi(n)$ such that: $\text{GCD}(e, \phi(n)) = 1$, i.e., e and $\phi(n)$ must be relatively prime.
5. Now compute the d , $1 < d < \phi(n)$ such that $e * d \equiv 1 \pmod{\phi(n)}$.

The public key exponent e is chosen, and the private key exponent d is calculated, so this (e, n) is the public key and (d, n) is the private key.

2.2 Encryption

Encryption of plaintext m is calculated by Eq. (1):

$$C = M^e \pmod{n} \quad (1)$$

where M is plaintext, n is modulus, cipher text is denoted as C , and e is encryption key exponent.

2.3 Decryption

Decryption requires modular exponentiation given by Eq. (2)

$$M = C^d \pmod{n} \quad (2)$$

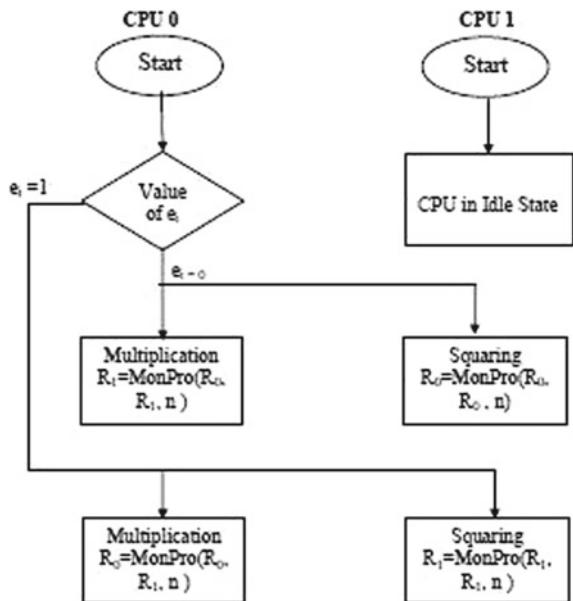
Literature suggests various methods for modular exponentiation: binary exponentiation, m-ary exponentiation, Montgomery powering ladder, m -ary powering ladders, etc. [7–11].

3 Proposed Approach for RSA Encryption

This section presents secure and efficient approach for RSA cryptosystem. Montgomery powering ladder is used for modular exponentiation. Compared to the binary exponentiation, it requires the 2's multiplication instead of 1.5 s [9], where s is the bits of exponent. But it prevents side-channel attack since it behaves very regularly. Each iteration computes two multiplications—modular squaring and modular multiplication whether the exponent bit is 1 or 0.

Algorithm 1 computes modular exponentiation using Montgomery powering ladder [9]. Here *MonPro* is the Montgomery modular multiplication method.

Fig. 1 Task division on the dual-core processors for i -th iteration



Algorithm 1: Parallel Montgomery powering ladder

- Function *MonPowLadder* (a, e, n)
- $R0 = 1 * r \bmod n$
- $R1 = a * r \bmod n$
- For $i = (s-1)$ to 0
//on processor 0
- $R[\sim ei] = \text{MonPro}(R0, R1, n)$
//on processor 1
- $R[ei] = \text{MonPro}(R1, R1, n)$
- Return $R0 = \text{MonPro}(R0, r \bmod n, n)$

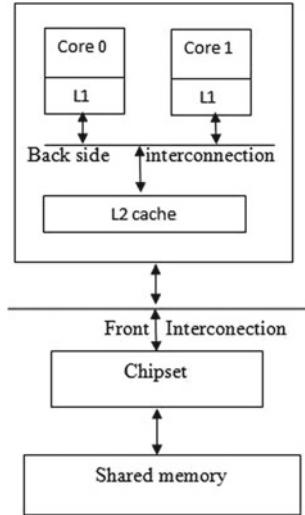
This method of parallelization had been proposed in [9], but no implementation results have been given.

There are two independent multiplications—modular squaring and modular multiplication for each iteration. Hence, the modular multiplication and modular squaring are assigned on the different cores of the process or independently as shown in Fig. 1. Therefore, it is referred as parallel Montgomery powering ladder in next section.

4 Implementation Results

RSA encryption is implemented on dual-core machine, comprising 2 intel (R) Core i3 CPU 2350 M 2.3 GHz dual-core processor and 4 GB RAM; the abstract view of

Fig. 2 Two cores sharing the arrays R0 and R1 in shared memory



the processors and RAM assembly has been shown in Fig. 2, with Python compiler version 2.7.615 running under Windows 7 Ultimate Service Pack 1 operating system.

Here multiprocessing module of Python is used for parallel implementation using two cores. Multiprocessing is a module in the Python that supports the parallel programming. It provides the process-level parallelism. It spawns the processes using the API similar to the process module, and for each processes, it creates an instance of the Python interpreter so that GIL could not interrupt the execution, that is how it prevents serialization of the parallel program [4, 5].

The Montgomery powering ladder is implemented with FIOS Montgomery modular multiplication where FIOS stands for finely integrated operand scanning [11–17] (Table 1).

To measure the enhancement in performance of the parallel Montgomery powering ladder, the comparison of results is done with the sequential version of the Montgomery powering ladder for various key sizes in range 1024–8192 bits.

Figure 3 depicts that parallel Montgomery powering ladder takes less time as compared to its sequential implementation.

Speedup [12–14] can be calculated by Eq. 3 as given below

$$\text{Speed-up} = \left(t_{\text{seq}} / t_{\text{par}} \right) \quad (3)$$

where t_{seq} and t_{par} are execution times for sequential and parallel Montgomery powering ladder. Theoretically, the parallel Montgomery powering ladder should achieve the speedup of factor 2, but the implementation results show the speedup of 1.63. This is due to synchronization overhead as the two cores are using data in shared memory. The efficiency of parallel Montgomery powering ladder as compared to its sequential implementation is 81%.

Table 1 Execution time for Montgomery powering ladder and parallel Montgomery powering ladder

Key size (in bits)	Montgomery powering ladder (in sec)	Parallel Montgomery powering ladder (in sec)
1024	0.8167	0.6534
1536	1.6499	1.0755
2048	2.5475	1.6516
2560	3.5271	2.19
3072	4.5831	3.7091
3548	5.6254	3.6857
4096	6.7374	4.2429
4608	8.2304	5.1147
5120	9.2868	5.7763
5632	11.0987	6.2964
6144	12.5159	7.4212
6656	14.3346	8.2953
7168	20.5004	9.475
7680	19.2532	10.4025
8192	21.3754	12.1651

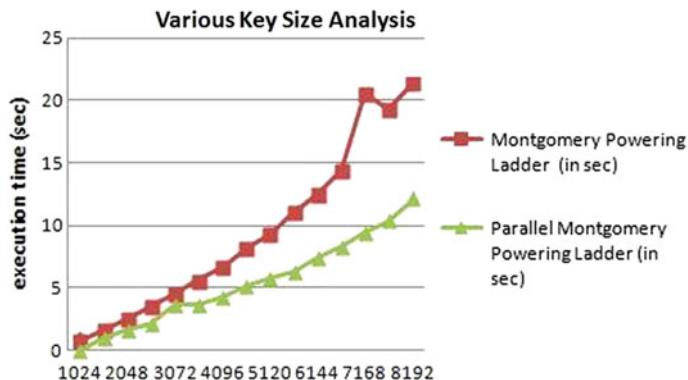


Fig. 3 Exponentiation performance w.r.t. various key sizes

5 Conclusion

The parallel Montgomery powering ladder is efficient when compared to its sequential implementation. This paper proposed parallel Montgomery powering ladder which is implemented on dual-core machine using the multiprocessing module of the Python. The performance can be improved further by identifying task parallelism and implementing on multicore systems.

References

1. Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and public key cryptosystems. *Commun. ACM* **21**(2), 120–126 (1978)
2. Selçuk, B., Sava, E.: Highly- parallel montgomery multiplication for multi-core general-purpose microprocessors
3. Dreslinski, B.G., Mudge, T.: A survey of multicore processors. *Signal Processing Magazine, IEEE* (2009)
4. Kumar, A., Ghrera, S.P., Tyagi, V.: An ID-based secure and flexible buyer-seller watermarking protocol for copyright protection, *Science & Technology. Pertanika J. Sci. Technol.* **25**(1), 57–76 (2017)
5. Kumar, A., Ghrera, S.P., Tyagi, V.: A comparison of buyer-seller watermarking protocol (BSWP) based on discrete cosine transform (DCT) and discrete wavelet transform (DWT). In: *Emerging ICT For Bridging The Future-Proceedings of The 49th Annual Convention Of The Computer Society Of India (CSI)*. vol. 1, pp. 401–408, Springer (2015)
6. William, S.: *Cryptography and Network security: Practices and principle* ISBN 10: 0-13-609704-9 ISBN 13: 978-0-13-609704-4
7. Cetin, K.K.: High-speed RSA implementation RSA laboratories RSA data security, Inc 100 Marine Parkway, Suite 500 Redwood City
8. Koc C.K., Acer, T.: Fast software exponentiation in GF (2^k), *IEEE Micro* (1997)
9. Joye, M., Yen, S.M.: The montgomery powering ladder. In: *Revised Papers from the 4th International Workshop on Cryptographic Hardwar and Embedded Systems, CHES '02*, pp 291–302, London, UK (2003)
10. Joye, M.: Highly regular m-ary powering ladders. *SAC2009, LNCS* **5867**, 350–363, Springer (2009)
11. Koc, C.K., Acar, T., Burton K.S.: Analyzing and comparing montgomery multiplication algorithms. *IEEE Micro* (1996)
12. Panda, M: Performance analysis of encryption algorithms for security. In: *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)* (2016)
13. Wang, Y., Leiwo, J., Srikanthan, T., Jianwen, L.: An efficient algorithm for DPA-resistant RSA. In: *APCCAS 2006—2006 IEEE Asia Pacific Conference on Circuits and Systems* (2006)
14. Kumar, A., Ansari, D., Ali, J., Kumar, K.: A new buyer-seller watermarking protocol with discrete cosine transform. In: *International Conference on Advances in Communication, Network, and Computing*, pp. 468–471. Springer Berlin Heidelberg
15. Kumar, A., Ghrera, S.P., Tyagi, V.: Modified buyer seller watermarking protocol based on discrete wavelet transform and principal component analysis. *Indian J. Sci. Technol.* **8**(35), 1–9
16. Kumar, A., Ghrera, S.P., Tyagi, V.: Implementation of wavelet based modified buyer-seller watermarking protocol. *WSEAS Trans. Signal Process.* **10**, 212–220
17. Kumar, A., Ghrera, S.P., Tyagi, V.: A new and efficient buyer-seller digital watermarking protocol using identity based technique for copyright protection. In: *Third International Conference on Image Information Processing (ICIIP)* IEEE

Validation and Optimization of Image Compression Algorithms



Nikhilesh Joshi and Tanuja Sarode

Abstract In current era, social media had been playing a major role considering fields like education, media, and politics. In such situation, information generated is an exponential form of which major chunk is in form of images. It becomes essential to use images compression techniques to save the storage space and bandwidth required for transmission. Image compression techniques help not only in reducing storage space requirement but improve the transmission bandwidth. In this article, an attempt to validate different compression algorithm for images using neural networks is highlighted. An attempt for optimization of compression algorithm is performed. Image compression algorithms like BTC, DCT, DWT, and SVD are considered, and optimization using the strength of each algorithm is performed. The results of optimization and validation technique show reduction in storage space using a variety of methodologies.

Keywords Image compression · Image quality metrics · Neural network · Block truncation coding · DCT · DWT · Backpropagation algorithm

1 Introduction

The availability of sophisticated user-friendly software and increased demand for multimedia-based technologies had made digital image data a necessity of our life. The amount of data generated everyday by a single user or business user is exponential in nature. The total information can be divided as 20% structured information while 80% unstructured. Images constitute the major part of information generated globally. Images require a huge chunk of space for storage and more transmission bandwidth.

N. Joshi (✉) · T. Sarode
Thadomal Shahani Engineering College (Affiliated, University of Mumbai) Bandra(W), Mumbai,
India
e-mail: nikhilesh.joshi@tsec.edu

T. Sarode
e-mail: tanuja.sarode@gmail.com

Image compression plays a vital role in reducing the storage space requirement and helps to increase transmission ratio over network. A picture/image/frame of size 256×256 consists of 65,536 pixels, and color image of size 640×480 has nearly a million pixels which needs storage chunk. If someone wants to download these files from the server or Internet, then it is time-consuming task. Data in the form of images comprise a significant portion of the multimedia. It also occupies the maximum part of the communication bandwidth required for multimedia transmission [1].

This enforces the development of efficient image compression to save storage space and minimize transmission bandwidth. The image compression technique most often used is transform coding. In transform coding compression techniques, initially frequency domain is used and then does its compression. The transform coefficient is always related to reduction in redundancy. The basic concept is to have a maximum amount of information stored in the smallest possible space [1, 2].

Redundancy and irrelevance are the two basic pillars of compression. The aim of removing duplicate information from the signal source (image/video) is termed as redundancy reduction. There are parts of signal which remain unnoticed by signal receiver that can be omitted termed as irrelevance reduction.

1.1 Performance Measure Metrics

The performance metrics are designed in order to compare the results of different image compression techniques. The performance metrics for image quality are broadly classified as subjective and objective. When there is a comparison regarding quality of image by evaluation of images and viewers read images, it is termed as subjective quality metrics. When statistical data-based indices are calculated to indicate image quality, it is termed as objective quality metrics. After reviewing the literature, it can be concluded that a certain set of parameters are involved in research to be performed. Some of the parameters can be standard parameter, while some of them may be applicable to few selected compression algorithms.

- Compression Ratio (CR):** The ratio of the size to original image to the size of compressed image is termed as compression ratio.

$$CR = \frac{\text{Size of Original Image}}{\text{Size of Compressed image}} \quad (1)$$

The performance of image compression can be specified in terms of compression efficiency which is measured by the compression ratio or by the bit rate.

- Peak Signal to Noise Ratio (PSNR):** It is qualitative measure which is based on mean square error (MSE) of the reconstructed image. It is shown in Eq. 2. The difference between original image and the reconstructed image is termed as MSE.

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [y(I, j) - x(I, j)]^2 \quad (2)$$

MSE helps to identify quality of reconstructed image using lossy compression technique. It is attractive measure for loss of image quality due to simplicity and mathematical convenience

The PSNR is calculated as the inverse of MSE. If the reconstructed image is close to the original image, MSE is small and PSNR takes large value. PSNR is dimensionless and expressed in decibel calculated as follows:

$$\text{PSNR} = 10 \log \left[\frac{L^2}{\text{MSE}} \right] \quad (3)$$

3. **Entropy:** It is an important factor to estimate whether the digital image is basically same as the original image. Entropy can be calculated by standard function available in MatLab. $E = \text{entropy}(I)$ returns E , a scalar value representing the entropy of grayscale image. Entropy is statistical measure of randomness that can be used to characterize the texture of the input image. It is defined as

$$E = -\sum(p * \log_2(p)) \quad (4)$$

where p contains the histogram counts returned from imhist .

4. **Correlation Measurement (C):** The similarity between two digital images could be quantified by correlation function. Each image is normalized by its root power, so the correlation measurement is defined as

$$\begin{aligned} C &= \frac{\sum_{m=1}^M \sum_{n=1}^N f(m, n) \bar{f}(m, n)}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N f^2(m, n) \sum_{m=1}^M \sum_{n=1}^N \bar{f}^2(m, n)}} \\ &= \frac{\sum_{m=1}^M \sum_{n=1}^N x(m, n) \bar{x}(m, n)}{\sqrt{\sum_{m=1}^M \sum_{n=1}^N x^2(m, n) \sum_{m=1}^M \sum_{n=1}^N \bar{x}^2(m, n)}} \end{aligned} \quad (5)$$

2 DCT, DWT, and SVD Method

When an image is represented as sum of sinusoids of varying magnitude and frequency, it is termed as discrete cosine transform (DCT) representation of an image. DCT property depicts that for an image; irrespective of whether its gray scale or color image, most of the visually significant information is concentrated in just a few coefficients of the DCT matrix. DCT divided image by separating it into the parts of different frequencies. The actual compression occurs in quantization step where the less important frequencies are discarded. The important frequencies are used

to retrieve the image in decomposition process, which results in distorted (reconstructed) image [3]

In most of the encoders which belong to mainstream, discrete cosine transform is used on large scale. DCT is mainly used to perform transform coding that maps time-domain signals to a frequency-domain representation. Frequency-domain spectrum can be compressed by truncating low-intensity regions. This will result in a large computation time, which will further double with increase in signal size. Henceforth, if one wants to calculate the DCT of the entire video frame, an unacceptable amount of time will be required. The Elucidation is to partition video into small blocks and applies DCT to each partition which may lead to degradation of picture quality. The discrete wavelet transform offers a better solution. DWT is a transform that maps time-domain signals to frequency-domain representations [2]. The DWT uses a set of digital filters which makes computation faster compared to DCT. This is possible because DWT is applied to entire signal and thus results in significant performance. The entire signal is analyzed by DWT resulting in better results. The DWT works as follows: It first separates those components of images which have a high frequency associated. Secondly, it resizes the remaining parts and finally rearranges them to form a new ‘transformed’ image [3].

SVD is stout and trustworthy orthogonal matrix decomposition method. Due to SVD intangible and constancy reasons, it becomes more and more popular in signal processing area. SVD is an eye-catching algebraic transform for image processing. SVD has conspicuous assets in imaging. In most image processing SVD properties are fully exploited; others still need more examination and subsidized. Many SVD properties are highly beneficial for images. Some of them can be listed as concentrated energy packing, explaining the least squares problem, and calculating pseudo-inverse of a matrix and multivariate analysis [2, 3]. A key property of SVD is its relation to the rank of a matrix and its ability to approximate matrices of a given rank. Digital images are often represented by low rank matrices and, therefore, able to be described by a sum of a relatively small set of eigen images. [3]

3 Optimized Image Compression Technique

In this section, we are focusing on modified block truncation algorithm, DWTDCT method, and single-value decomposition method.

3.1 Enhanced DCTBTC Algorithm

In block truncation coding algorithm, there was a problem of staircase at edges in the reconstructed image. In the BTC algorithm, different blocks are selected depending on requirement. For experimental purpose, we will have N as block size which

can be 2/4/8/16/32. Depending on the block size, mean and standard deviation are calculated. The compressed bit map is obtained by

$$B = \begin{cases} 1 & w_i > \mu \\ 0 & w_i \leq \mu \end{cases} \quad (6)$$

The bitmap B mean and standard deviation are transmitted at the decoder end. Next, H and L are calculated as highlighted in Eqs. 7 and 8, and reconstructed image can be obtained by replacing the element 1 in the B with H value and element 0 with L. The reconstructed highlighted in Table 1. In this process, the parameters like MSE, PSNR, correlation, and CR are calculated and highlighted in Table 1. The image size varies from 512×512 and 256×256 .

$$H = \mu + \sigma \sqrt{\frac{p}{q}} \quad (7)$$

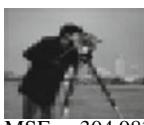
$$L = \mu - \sigma \sqrt{\frac{p}{q}} \quad (8)$$

p Number of 0's in bit map

q Number of 1's in bit map

In the proposed algorithm which is termed as discrete cosine transform block code truncation (DCTBTC), DCT is applied to the image and after that BTC is applied.

Table 1 Comparative result of BTC & DCTBTC

Block size 8		Block size 16	
BTC	DCTBTC	BTC	DCTBTC
			
MSE = 406.4163 RME = 20.1598 PSNR = 50.8298 CR = 6.4 Corr = 0.9468	MSE = 304.9837 RME = 17.4638 PSNR = 53.7010 CR = 6.4 Corr = 0.9563	MSE = 567.2139 RME = 23.8163 PSNR = 47.4962 CR = 7.52 Corr = 0.9257	MSE = 458.5847 RME = 21.4146 PSNR = 49.6221 CR = 7.52 Corr = 0.9341
			
MSE = 173.59 RME = 13.17 PSNR = 59.2581 CR = 6.4 Corr = 0.9615	MSE = 148.2788 RME = 12.1770 PSNR = 60.9126 CR = 6.4 Corr = 0.9659	MSE = 318.87 RME = 17.84 PSNR = 53.18 CR = 7.52 Corr = 0.9285	MSE = 286.4320 RME = 16.9243 PSNR = 54.3285 CR = 7.52 Corr = 0.9333

One important change is that we first perform low-pass filtering on the image, and then, DCTBTC is applied. In this way, we can achieve a better image quality and correlation can be enhanced at the cost of MSE and PSNR which is tabulated in Table 1

If we have 256×256 images and block size 4×4 , then the number of bits required is 524,288 bits. In compressed form, it is represented by 0 or 1, so only 1 bit is required, so 4×4 block requires 16-bit representation and L and H which are transmitted require 16 bits, so only 32 bits are required. In this way, the total bits required for transmission in compressed form will be $64 \times 64 \times 32 = 131072$ bits. So it can be observed that transmission requires a less number of bits which result in less requirement of bandwidth and reduce storage space. The compression ratio can be calculated as given by Eq. (1); here, it comes out to be 4; that is, 25% compression can be achieved by this method.

3.2 DWTDCT Algorithm and Neural Network Validation

In this section, we tried to implement DWTDCT method on the a set of images. The proposed algorithm is as follows

- Step 1: Load the image to be compressed.
- Step 2: Split the original image to Y, C_b, and C_r color planes.
- Step 3: Decompress the image planes using DWT.
- Step 4: Apply sub-band coding and shift data to create zero matrix.
- Step 5: Create the transform array using DCT and eliminate the zero matrixes using block N × N.
- Step 6: Reconstruct the image using iDCT. Save the compressed image, and calculate the compression performance parameter.

The neural network architecture consists of 4096 input neurons since the input image is of size $64 \times 64 = 4096$ pixels. Each neuron will represent one intensity value. There are 80 hidden layer neurons. Since we want the compression ratio for the range 10–90%, there are nine output neurons, each corresponding to the appropriate compression ratio. The architecture is as shown in Table 1. The neural network has three set of input:

Training set: It consists of 30 images for training.

Testing set: It consists of 20 images for testing the NN.

Unknown set: It consists of 10 images for checking NN.

The NN is trained using backpropagation algorithm, and then, testing is done using testing set, and results are noted. The correlation between compressed image and original image is calculated. It is observed that most of the result match and the once trained neural network give output in less than 0.003 s

The performance of the method was evaluated and tabulated as shown in Tables 2 and 3. Table 2 highlights the DWTDCT method with performance parameters. Table 3

Table 2 DWT-DCT method

Original image	Compressed image	Parameters
		Original size = 65 KB Compressed size = 17 KB MSE = 177.07 RME = 133.305 SNR = 167.315 PSNR = 13.050 Compression = 75%
		Original size = 13.3 KB Compressed Size = 4.95 KB MSE = 164.22 RME = 128.149 SNR = 168.388 PSNR = 13.839
		Original size = 8.42 KB Compressed Size = 2.46 KB MSE = 326.26 RME = 180.626 SNR = 165.017 PSNR = 6.974 Compression = 70.78%

highlights the correlation and entropy used for training image. Table 3 indicates entropy and correlation parameter. Correlation indicates the matching between original image and reconstructed image, i.e., compressed image. In most of the images that are tested for optimization, it is observed that more than 99% image is reconstructed after applying compression technique. The table indicates values for correlation for various test images in last column. Entropy is obtained in three stages for given set of test image. When we calculate entropy, the compressed value at each level should not exceed the value obtained for the original image. For all images, this pattern prevails; thus, it justifies that optimization is obtained. It is observed that for the same set of images, single-value decomposition yields the same results as reflected in Table 4. The table reflects the entropy value of original image followed by SVD value. In all three methodologies, neural network was designed and trained using backpropagation algorithm. The network was trained and yielded the same result for all set of images. The purpose of using neural network was to validate the methodology implemented.

Table 3 Entropy and correlation for DWT-DCT

Image	Original	DWT	DCT	Correlation
Test1.jpg	7.128	6.2112	3.4781	0.9966
Test2.jpg	7.6405	5.7263	3.7794	0.9932
Test5.jpg	7.4287	4.6333	3.4684	0.9959
Test9.jpg	7.431	7.6511	3.4417	0.9935
Test11.jpg	6.7616	6.8725	3.8558	0.9726
Test15.jpg	7.6317	6.6027	3.8717	0.9919
Test18.jpg	7.4908	5.3608	3.9029	0.9854
Test20.jpg	4.1354	3.5140	3.8557	0.9908
Test24.jpg	7.4884	5.1169	3.7457	0.9719
Test30.jpg	6.1985	6.1706	3.6054	0.9909

Table 4 Entropy and correlation for SVD

Image	Original	SVD	Correlation
Test1.jpg	7.79	7.80	0.991
Test2.jpg	7.64	7.45	0.995
Test5.jpg	7.42	7.12	0.996
Test9.jpg	7.12	7.05	0.991
Test11.jpg	6.76	6.42	0.994
Test15.jpg	7.61	7.52	0.910
Test18.jpg	7.43	7.24	0.995
Test20.jpg	4.15	4.05	0.997
Test24.jpg	6.81	6.70	0.998
Test30.jpg	5.98	5.90	0.993

4 Conclusion

In this paper, methods are suggested for optimization of image compression algorithm. It can be concluded that result obtained indicates the reduction in storage space requirement. This will directly help in reducing the transmission bandwidth requirement for various images. The design of NN helps in reducing the time requirement for calculating the compression ratio. The neural network also helps in validating the result obtained before training by comparing result obtained after training. The limitation of our method is that currently it deals with grayscale images. The neural network training consumes a lot of time and requires adjustment of parameter.

References

1. Gonalez, R.C., Woods, R.E.: Digital Image Compression, 3rd edn, Prentice Hall
2. Delp, E.J., Mitchell, O.R.: Image compression using block truncation coding. *IEEE Trans. Commun.* **27**(9), 1335–1342 (1979)
3. Lema, M.D., Mitchell, O.R.: Absolute moment block truncation coding and its application to color images. *IEEE Trans. Commun.* **COM-32**(10), 1148–1157 (1984)
4. Cheng, S.C., Tsai, W.S.: Image compression by moment-preserving edge detection. *Pattern Recogn.* **27**(11), 1439–1449
5. Desai, U.Y., Muzuki, M.M., Horn, B.K.P.: Edge and mean based compression. MIT Artificial Intelligence Laboratory AI Memo No.1584 (1996)
6. Ammarunnishad, T.M., Govindan, V.K., Abraham, T.M.: Improving BTC image compression using a fuzzy complement edge operator. *Sig. Process. Lett.* **88**(12), 2989–2997 (2008)
7. Ammarunnishad, T.M., Govindan, V.K., Abraham, T.M.: A fuzzy complement edge operator. In: IEEE Proceeding of the Fourteen International Conference on Advance Computing and Communication Mangalore, Karnataka, India (2006)
8. Kumar, A., Singh, P.: Futuristic algorithm for gray scale image based on enhanced block truncation coding. *Int. J. Comput. Inf. Sys.* **2**(5), 53–60 (2011)
9. Mathews, J., Nair, M.S., Jo, L.: Modified BTC algorithm for gray scale images using max-min quantizer. *IEEE Trans.* 377–382 (2013)
10. Nadenau, M.J., Reichel, J., Kunt, M.: Wavelet Based color image compression: exploiting the contrast sensitivity function. *IEEE Trans. Image Process.* **12**(1), 58–70 (2003)
11. Gupta, M., Garg, A.K.: Analysis of image compression algorithm using DCT. *Int. J. Eng. Res. Appl.* **2**(1), 515–521 (2012). ISSN: 2248–9662
12. Singh, M., Garg, M.: Mixed DWT-DCT approached based image compression technique. *Int. J. Eng. Comput. Sci.* **3**(11), 9107–9111 (2014). ISSN:2319–7242
13. Sagwan, B., Sharma, M., Gupta, K.: RGB based KMB image compression technique. In: International Conference on Reliability, Optimization and Information Technology (2014)
14. Gupta, M., Garg, A.K.: Analysis of image compression algorithm using DCT. *Int. J. Eng. Res. Appl.* **2**(1), 515–521 (2012). ISSN: 2248–9662
15. Singh, M., Garg, M.: Mixed DWT-DCT approached based image compression technique. *Int. J. Eng. Comput. Sci.* **3**(11), 9107–9111 (2014). ISSN:2319–7242
16. Sagwan, B., Sharma, M., Gupta, K.: RGB based KMB image compression technique. In: International Conference on Reliability, Optimization and Information Technology (2014)
17. Singh, M., Garg, M.: Mixed DWT DCT approached based image compression technique. *Int. J. Eng. Comput. Sci.* **3**(11), 9008–9111 (2014). ISSN 2319-7242

Temperature Control Using Sliding Mode Control: An Experimental Approach



Krupa Narwekar and V. A. Shah

Abstract In this work, the variable structure sliding mode control is applied to the temperature control of the reactor in the laboratory. The sliding mode control law is achieved using the conventional sliding mode control. This result is compared with the modified reaching law in sliding mode control, and the results are observed and compared experimentally. Hence, tracking problem is handled using sliding mode control practically.

Keywords Temperature control · Reactor · Sliding mode control · Non-linear · Power rate reaching law

1 Introduction

Temperature control is an important parameter in the industry, which needs to be controlled. In most of the reactors, the final product quality is actually a function of temperature in the reactor [1]. Therefore, reactor temperature control is considered in this work. The temperature control is achieved in the most simplest form which is by PID control techniques [2]. Due to the development of the advanced control and simulation softwares, new control techniques have been developed by the researchers. The application of these techniques is found in the literature for controlling temperature. In [3], fuzzy control is used to control temperature. Also, neural network is used to control temperature [4–7].

In last 5–6 decades, a lot of development is there in variable structure control techniques. In variable structure control, the controller has discontinuous movement to achieve the control [8]. The variable structure sliding mode control is found in the literature with a large number of applications in various areas like process industry,

K. Narwekar (✉)
SVIT, Vasad, affiliated to GTU, Ahmedabad, Gujarat, India
e-mail: krupadn@gmail.com

V. A. Shah
Dharamsinh Desai University, Nadiad, Gujarat, India

chemical industry, robotics applications, in ac-to-dc converters. To mention a few, in [9], SMC is used to control glucose feed-in-feed batch fermentation process. It is also used to control speed of DC motor [10]. In [11], temperature control of batch reactor is achieved using SMC. The sliding mode control is a robust control which consists of two steps to achieve the control. One is designing the sliding surface suitable for application and another selecting the reaching law for the states to reach the sliding surface. Once the states reach the sliding surface, the controller becomes robust to match disturbances and parametric uncertainties [8]. Because of these properties of SMC, it is used to control various parameters as mentioned previously in this paragraph. The main work in this paper is real-time application of sliding mode control to maintain the temperature of the reactor. Therefore, a tracking problem is considered. The temperature is reached to the desired level and is maintained even in the presence of matched disturbance. The classical SMC is first applied to control the temperature, and then, a modification, i.e. power rate reaching law, is used to control the temperature. Both the results are compared and results are observed. The paper is organized as follows. In Sect. 2, sliding mode and its theory are explained in detail. In Sect. 3, the experimental setup is discussed. In Sect. 4, design of the control law is discussed, and then, Sect. 5 is results and discussions followed by Sect. 6 conclusion.

2 Preliminaries of Sliding Mode Control

The sliding mode control in its simplest form consists of designing the sliding surface and the reaching phase. The surface can be a linear, a plane surface or nonlinear depending on the equation of the surface. As discussed in introduction variable structure control, sliding mode control is robust controller and has been applied to numerous systems [12–15], etc. Last five decades, there is rigorous research carried on sliding mode theory. The design of SMC begins with selecting a proper sliding surface and a reaching phase to ensure the states to reach the designed sliding surface. However, the system is sensitive to disturbances in the reaching phase. Hence to reduce the reaching time or completely eliminating the reaching phase can be the solution to reach the sliding surface as early as possible. Eliminating the reaching phase the systems becomes highly sensitive and may become unstable [16]. Therefore, the efforts for reducing the reaching phase is carried out by several authors for many systems. The basics of sliding mode control are as follows.

Let us consider a second-order process given in

$$\begin{aligned}\dot{x}_1 &= x_2 \\ \dot{x}_2 &= a(x) + b(x)u + d\end{aligned}\tag{1}$$

where $[x_1 x_2]^T$ is the state vector $a(x)$, $b(x)$ nonlinear function in x , u being the control input. d is the matched disturbance.

Sliding surface is designed as

$$s = c_1x_1 + c_2x_2 \quad (2)$$

For the system to be on equilibrium or desired position, the state trajectories should reach the sliding surface. Therefore, designing the sliding surface includes finding the values of c_1 and c_2 . These can be found out by pole placement or LQR techniques for linear systems. Reaching the sliding surface implies $s = 0$; therefore, $\dot{s} = 0$

so

$$s\dot{s} = \eta|s| \quad (3)$$

where $\eta > 0$ ensures finite-time convergence to $s = 0$, and it is also called η reachability condition. It is clear that the states move from initial condition to the sliding surface based on this reaching law equation. It should also be noted that when the states are in this phase, they are sensitive to process uncertainty and disturbances, so it is clear that the reaching phase should be as small as possible. Many researchers are striving to minimize the reaching phase by increasing the gain during this phase, but it has been observed that due to it, there are large oscillations and system becomes extremely sensitive to the unmodelled dynamics [17–20]. Therefore, our aim is to implement the reaching law which has fast convergence, so that the temperature tracking of batch reactor is handled in efficient manner. The various reaching laws are proposed by Gaos, Hungs, and Bartoszewicz [17] in the literature of sliding mode control.

Constant Rate Reaching Law

$$\dot{s} = -k sign(s) \quad (4)$$

Proportionate Reaching Law

$$\dot{s} = -(k_1 + k_2|s|)sign(s) \quad (5)$$

Power Rate Reaching Law

$$\dot{s} = -k|s|^\alpha sign(s) \quad 0 < \alpha < 1 \quad (6)$$

3 The Experimental Setup

As discussed in Sect. 1, there are lot of simulation studies done on procprocess and chemical ess parameters in different applications. Motivated by [11], the temperature control using SMC of laboratory reactor is carried out in this work. The process trainer, CE117 of TecQuipment, is used for experimental validation of SMC algorithm. Figure 1 shows the CE117 Process Trainer.



Fig. 1 Experimental setup

The working of the reactor is as follows: the hot water from the heater is passed through the coil in the reactor vessel. The tank water is circulated through the condenser unit into the tank. The interfacing of this kit with the computer system is done through DAQ NI6009 card of National Instruments. The design of the controller is done in the LabVIEW environment. The temperature sensor (RTD) and transmitter output are taken through the DAQ card to the LabVIEW, the control law is designed using sliding mode control law, and the output from the LabVIEW through the DAQ card is given to the pump for manipulating the hot water flow through the coil.

4 Design of the Control Law

As seen in the previous section, the experimental setup consists of water in the tank whose temperature is to be controlled. To design the control law, it is necessary to model the system. By using mass balance and energy balance equation, we can write dynamic model of the reactor as

$$\frac{dT}{dt} = \frac{W_i}{V\rho}(T_i - T) + \rho q_h(T_h - T_0) \quad (7)$$

where

T = temperature of the water in the tank ($^{\circ}\text{C}$)

W_i = inlet mass flow rate (kg/s)

V = volume of fluid in the process vessel (L)

ρ = fluid density (kg/m^3)

T_i = inflow temperature ($^{\circ}\text{C}$)

T_h = temperature in the heater tank ($^{\circ}\text{C}$)

T_0 = outlet temperature from the heater tank ($^{\circ}\text{C}$)

q_h = flow rate through the heat exchanger (L/min).

By varying the flow rate (q_h) and outlet of heat exchanger, the temperature of the tank is manipulated. For designing the control law using sliding mode control, let us select the sliding surface as

$$s = T_d - T \quad (8)$$

where T_d is the desired temperature. As discussed in Sect. 2, reaching the sliding surface is

$$s = 0 \quad (9)$$

therefore

$$\dot{s} = 0 \quad (10)$$

by using (4) and (10)

$$\dot{T}_d - \dot{T} = -k\text{sign}(s) \quad (11)$$

Substituting (7) in (11) and finding the equivalent control law for q_h ,

$$q_h = \frac{8380}{(T_h - T_0)} [\dot{T}_d - 2.22e^{-5}(T_i - T) + k\text{sign}(s)] \quad (12)$$

(12) gives the control effort required to keep the temperature at desired level. Again by modifying the reaching law, the reaching time is improved as well as the chattering is reduced to some extent. (13) is the control effort with power rate reaching law. The reactor is given the input through the pump to as per the flow rate generated by (12) and (13).

$$q_h = \frac{8380}{(T_h - T_0)} [\dot{T}_d - 2.22e^{-5}(T_i - T) + k|s|^{\alpha}\text{sign}(s)] \quad (13)$$

5 Results & Discussion

The results are obtained by using the control law derived in the previous section for conventional SMC as well as modified SMC. The results obtained are showing the advantage of using power rate reaching law. Figure 2 shows the temperature tracking

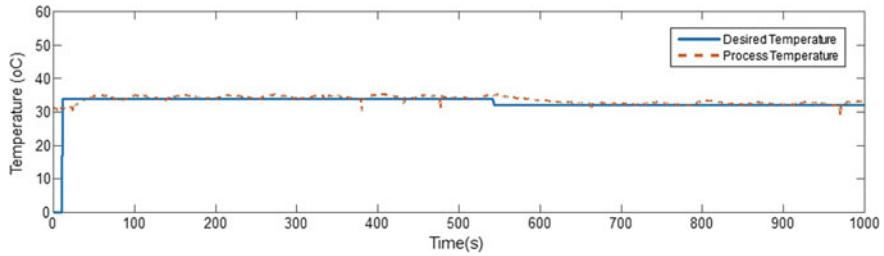


Fig. 2 Temperature tracking using SMC

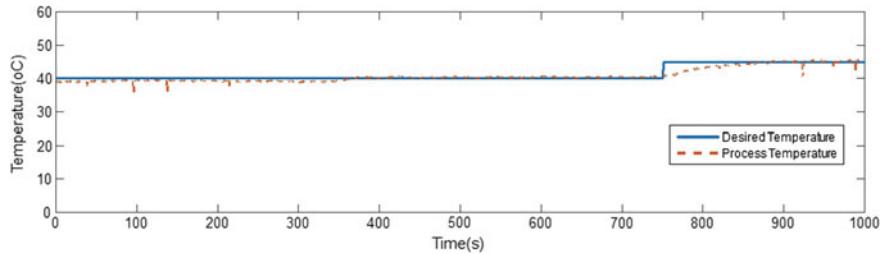


Fig. 3 Temperature tracking using modified SMC

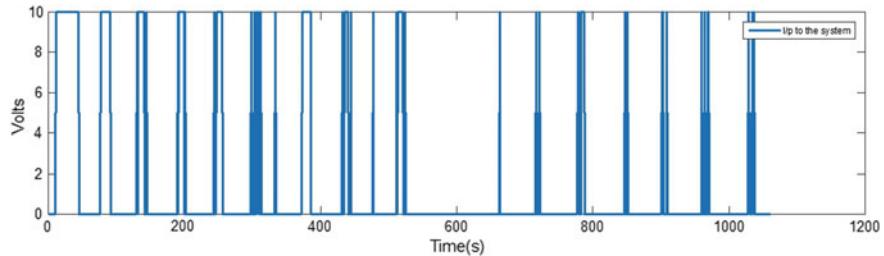


Fig. 4 Control law with SMC

using conventional SMC. Figure 4 shows the control efforts to achieve the desired temperature. Figure 3 shows the temperature tracking for modified reaching law which shows better tracking with respect to Fig. 2. The control efforts for achieving the desired temperature using modified SMC are observed in Fig. 5. However, we can observe that the chattering is reduced in Fig. 5. The error curve as shown in Fig. 6 in shows the improvement in tracking by modified SMC as compared to conventional SMC. The tuning parameter, i.e. gain k , is set to 15, and α is set to 0.7 to achieve the desired performance.

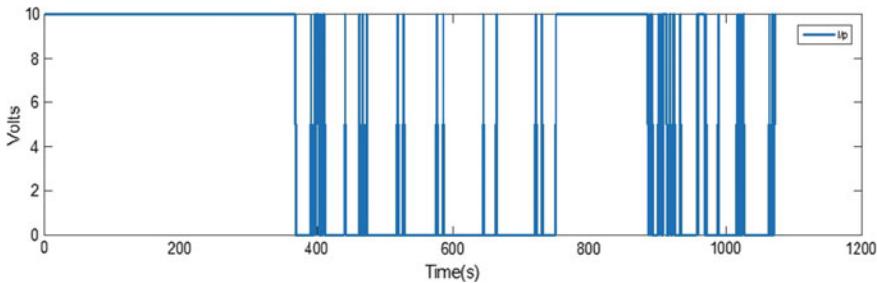


Fig. 5 Control law with modified SMC

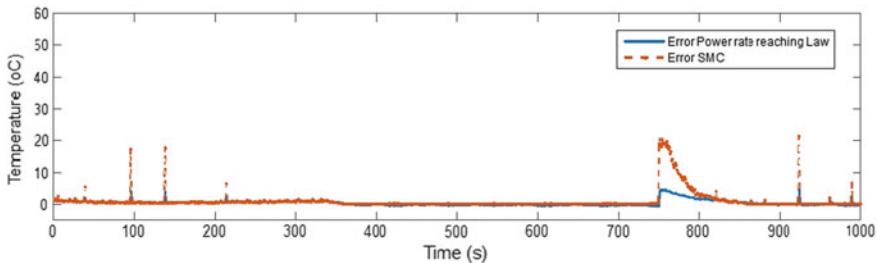


Fig. 6 Error curve

6 Conclusion

Thus, the temperature control of the laboratory reactor is achieved in real time. The design and application of the control law in real-time setup, which has nonlinearities, dead zones, friction, etc., are though challenging. In this work, it is achieved successfully. Also, the effectiveness of the power rate reaching law over the conventional SMC is validated experimentally. The work can be extended to higher-order sliding mode control application. As the laboratory reactor system in this work is commonly used in many processes and chemical industries, so control strategies can be tested and applied, and the results can be validated.

Acknowledgements The authors thank Pandit Deendayal Petroleum University, Gandhinagar, Gujarat, India, for permitting us to use experimental setup.

References

1. Rotstein, G.E., Lewin, D.R.: Control of an unstable batch chemical Reactor. *Comput. Chem. Eng.* **16**(1) (1992) 27–49 (2001)
2. Ranjan, S., Sharma, A., Chaudhary, P.: An effective temperature controller system using PID mechanism. In: 2014 Innovative Applications of Computational Intelligence on Power, Energy

- and Controls with their Impact on Humanity (CIPECH), IEEE, Ghaziabad, India, 28–29 Nov 2014 <https://doi.org/10.1109/CIPECH.2014.7019086>
- 3. Pamela, D., Jebarajan, T.: Intelligent controller for temperature process. *Int. J. Control Automat.* **6**(5), 191–198 (2013). <https://doi.org/10.14257/ijca.2013.6.5.18>
 - 4. El Kebir, A., Chaker, A., Negadi, K.: A neural network controller for a temperature control electrical furnace. *Int. Rev. Autom. Control (IREACO)*, **6**(6), 689–694 (2013)
 - 5. Chen, L., Ge, B., de Almeida, A.T.: Self-tuning PID temperature controller based on flexible neural network. In: International Symposium on Neural Networks ISNN 2007: Advances in Neural Networks ISNN 2007, pp. 138–147 (2007)
 - 6. Cui, X., Shin, K.G.: Application of neural networks to temperature control in thermal power plants. *Eng. Appl. Artif. Intell.* **5**(6), 527–538 (1992). <http://hdl.handle.net/2027.42/29760>
 - 7. Parkale, Y.V.: Comparison of ANN controller and PID controller for industrial water bath temperature control system using Matlab Environment. *Int. J. Comput. Appl.* (0975–8887), **53**(2) (2012)
 - 8. Utkin, V.I.: Variable structure systems with sliding modes. *IEEE Trans. Auto. Control* AC-22(2) (1977)
 - 9. Georgieva, O., Hristozov, I., Pencheva, T., Tzonkov, St., Hitzmann, B.: Mathematical modelling and variable structure control systems for fed-Batch fermentation of escherichia coli. *Chem. Biochem. Eng. Q.* **17**(4), 293–299 (2003)
 - 10. Rhif, A.: Stabilizing Sliding mode control design and application for a DC motor: speed control. *Int. J. Instrumentation Control Syst. (IJICS)* **2**(1) (2012)
 - 11. Chen, C-T.: A sliding mode control strategy for temperature trajectory tracking in batch processes. *IFAC Proc.* **45**(15), 644–649 (2012)
 - 12. Sira-Ramirez, H.: Variable structure control for nonlinear system. *Int. J. Syst. Sci.* **18**(9), 1673–1689 (1987)
 - 13. Piltan, F., Sulaiman, N., Rashidi, M., Tajpeikar, Z., Ferdosali, P.: Design and implementation of sliding mode algorithm: applied to robot manipulator—a review. *Int. J. Rob. Autom. (IJRA)*, **2**(5) (2011)
 - 14. Kurode, S., Bandyopadhyay, B., Gandhi, P.S.: Robust control for slosh-free motion using sliding modes. In: 12th international workshop on variable structure systems (VSS), 440–445 (2012)
 - 15. Mujumdar, A., Kurode, S., Tamhane, B.: Fractional order sliding mode control for single link flexible manipulator. In: IEEE International conference on control applications (CCA), pp. 288–293 (2013)
 - 16. Mehta, A., Bandyopadhyay, B.: Frequency-shaped and observer-based discrete-time sliding mode control. Springer Briefs in Applied Science Technology (2015)
 - 17. Bartoszewicz, A.: A new reaching law for sliding mode control of continuous time systems with constraints. *Trans. Instit. Measur. Control* **37**(4), 515–521 (2015)
 - 18. Zhang, M., Yu, Z., Zhou, Y., Huan, H.: The sliding mode variable structure control based on composite reaching law of active magnetic bearing. *CIC Express Lett. ICIC Int.* 2008 ISSN 1881-803X **2**(1), pp. 59–63 (2008)
 - 19. Ma, M., Juiping, B., Chen, M.: Improved Reaching Law Sliding Mode Control Algorithm Design for DC Motor Based on Kalman Filter. *Int. J. Signal Process. Image Process. Pattern Recog.* **8**(4), 349–360 (2015)
 - 20. Koshy, R.A., Thomas, S.: A two-surface discrete sliding mode control based on approach angle reaching law. In: Proceedings of the international Multi Conference of Engineers and computer scientists 2012, Hongkong, March 14–16 (2012)

Simulation Platforms for Wireless Sensor Networks: How to Select?



Aarti Kochhar, Pardeep Kaur and Preeti

Abstract With the fast-track growth of IoT (Internet of Things), wireless sensor networks (WSN) has also gained attention. For effective communication among sensor nodes, protocols are being developed rapidly. Simulation is necessary for any network before deployment to check the functionality and detect the potential errors. Earlier recognition of issues can save both resources and time of network designers. Few simulation environments were already available before the development of WSN technology. Later on, a few other simulation environments were developed specifically for WSN. Simulation environments based on programming language C/C++ and MATLAB are reviewed in this paper. This paper also discusses the extensions of simulation environments. The main focus of this paper is to help novice researchers in choosing an appropriate simulation tool that fits their research criterion.

Keywords Wireless sensor network · Simulation tools · MATLAB · Protocol

1 Introduction

Simulation environment models the expected behavior of a network in a real-life scenario. Simulation environments or tools use various testing and analyzing techniques to evaluate the network. Different researchers have categorized the tools depending upon different criteria [1, 2]. This paper categorizes and reviews the simulation environments present in the literature on the basis of the scripting language of the tool. Since there is a wide range of simulation tools available, this paper presents a survey of only the most widely used simulation tools. Section 2 discusses simulation tools based on C/C++, and Sect. 3 discusses simulation tools based on MATLAB. The paper is concluded in Sect. 4 in the form of a comparison table.

A. Kochhar (✉)

Lovely Professional University, Phagwara, India

e-mail: artikochhar92@gmail.com

P. Kaur · Preeti

University Institute of Engineering and Technology, Panjab University, Chandigarh, India

2 Simulation Environments Based on C/C++

2.1 *Ns-2*

Originally NS (network simulator) was developed for simulation of routing and multicast protocols. Later on, it was upgraded to support WSNs. NS-2 is an open-source event-driven simulator written in object-oriented C++ and OTcl. Since NS-2 is an open-source simulation tool, hence it saves the cost of simulations. NS-2 was originally designed for IP networks, so it does not have any unique characteristics for WSN but NS-2.30 has few inbuilt packages for underwater WSN. Although NS-2 can support only two MAC (medium access control) protocols, i.e., TDMA (Time Division Multiple Access) and IEEE 802.11, but support of IEEE 802.15.4 is available in version 2.28. It also supports few popular routing protocols like DSDV (Destination Sequenced Distance Vector), DSR (Dynamic Source Routing) and AODV (Adhoc On-demand Distance Vector). It also supports multicasting through announcements. Energy model available in NS-2 is not very reliable and accurate because it shows linear depletion of energy and a fixed starting value which is not always a case in real-life WSN scenarios. Trace files obtained after simulating for larger networks are very tedious and hard to manage. Since NS-2 allows very less customization, advanced learning is required to perform simulations [3].

Mannasim framework of NS provides LEACH (Low Energy Adaptive Clustering Hierarchy) and directed diffusion [4]. It also provides Bluetooth, IR-UWB (Impulse Radio Ultra-Wide Band), IEEE 802.16. SensorSim [5] based on NS-2 provides support of lightweight protocols to microsensors.

2.2 *OMNeT ++/Castalia*

OMNeT++ (Object Modular Network Testbed in C++) is discrete event-driven simulation environment [6]. It is based on object-oriented framework. It is not a simulator but provides frameworks to write simulations for WSNs. OMNeT++ supports Windows and Linux. Mobility framework supports MAC layer and mobility. Mainly, OMNeT++ is for the support of IP (Internet Protocol) networks, so it does not have specific protocols and energy models for WSNs. Also, the tool can capture delay time for lower layers of protocol stack, but it is not able to capture the delay for the application layer.

Castalia is OMNeT++ based simulator for wireless body area network [7]. Castalia has built-in sensing devices for pressure, temperature and light. It also provides a flexible base of MAC protocols. Limitation of Castalia is that it provides only generic framework for validation of code. It cannot validate codes for specific applications. INET, a protocol library in OMNeT++, supports routing protocols.

2.3 GloMoSim/QualNet

GloMoSim (Global Mobile Information System Simulator) is discrete simulation model based on C for sequential and parallel execution. There were no default packages for WSN. In 2000, QualNet was developed as a derivate of GloMoSim [8]. No further improvement was made in GloMoSim. QualNet was enhanced to provide security and Zigbee support. Later on, a GUI (Graphical User Interface) and a visualizer were added for 2D and 3D view. QualNet allows the user to create new protocol models. Also, it enables the user to optimize and analyze existing models. One can design a heterogeneous network also on QualNet. It supports both Cartesian and geographical coordinate system. It also provides few inbuilt channel models, propagation models and mobility model. Major limitation of the simulation tool is that it has limited number of inbuilt models. For example, it supports only one mobility model, i.e., random waypoint mobility model.

2.4 TOSSIM

TOSSIM (TinyOS Simulator) is discrete event simulator for TinyOS' WSN which captures the behavior of motes [9]. Motes are TinyOS sensor nodes. TOSSIM is simple and efficient simulator which is based on nesC (Network Embedded Systems C). It uses abstraction to capture hidden terminal problem, data corruption and system failures. It also provides services to communicate with external applications. TOSSIM treats each hardware resource as a component as it has a component-based programming model. It replaces each component with its simulated implementation. TOSSIM can simulate a protocol till bit-level. There are few limitations of TOSSIM such as: Execution model of TOSSIM does not capture CPU time. Also, it assumes that each node in the network runs the same code which is not true in case of heterogeneous networks. TOSSIM does not provide any energy consumption model. Although installation of TOSSIM on Linux platform is easy, on Windows platform installation has few prerequisites that increase the complexity of process.

TOSSIM has provision to add link between source and destination. Gain and noise can be added to the link. Mean and standard deviation of the noise floor can be changed. Sensitivity of the receiver mote can also be defined. Although parameters like symbols per second and preamble length have a default value already set, it can be changed by using #define directive, similar to C languages.

3 Simulation Environments Based on MATLAB

3.1 MATLAB/Simulink

MATLAB stands for matrix laboratory. MATLAB provides an interactive model, a number of built-in mathematical functions and advanced help options. It also provides user-defined functions. Simulink provides a graphical user interface for constructing block diagrams. Simulink has a large library of blocks. User can customize the parameters of these blocks. Block diagram is finally simulated for the results. These simulation results can be used in MATLAB for further processing [10]. Other than Simulink, ThingSpeak is also supported by MATLAB. ThingSpeak [11] is an open-source application to collect data from nodes over HTTP (Hypertext Transfer Protocol) using the Internet. Collected data can be analyzed using MATLAB.

3.2 PROWLER/RMASE

PROWLER is probabilistic wireless network simulator [12]. It is an event-driven simulator which was developed for simulation of MICA motes' network. It can either operate in probabilistic mode or in deterministic mode. All the codes in PROWLER are written in MATLAB with .m extension. PROWLER has a graphical user interface; hence, it is user-friendly. User can see the time of collision along with the ID of mote where collision occurred. Figure 1 shows the user interface of the simulation tool. Red dots show the transmitting motes, green dots represent receiving motes, and yellow dots represent receiving of collided packet. Motes can be set up in various other topologies also. It also supports few radio propagation models which help in determining fading effects. Drawback of PROWLER is that it supports only one MAC protocol, i.e., IEEE 802.11.

RMASE is an extension of PROWLER built specifically for routing protocols. RMASE has provision of defining behavior of source and destination. To compare different routing protocols, RMASE has predefined performance metrics like energy consumption, throughput and latency.

3.3 HarvWSNet

Most of the available simulation environments model only linear battery which cannot be used to accurately estimate the lifetime of EH-WSN (Energy Harvesting Wireless Sensor Network). So, a separate simulator is required for the modeling of EH-WSN. HarvWSNet is a framework suited for EH-WSN [13]. It helps in predicting the lifetime of network. It is a co-simulation framework which combines the capabilities of two tools: MATLAB and WSNet. MATLAB helps in modeling energy harvesting

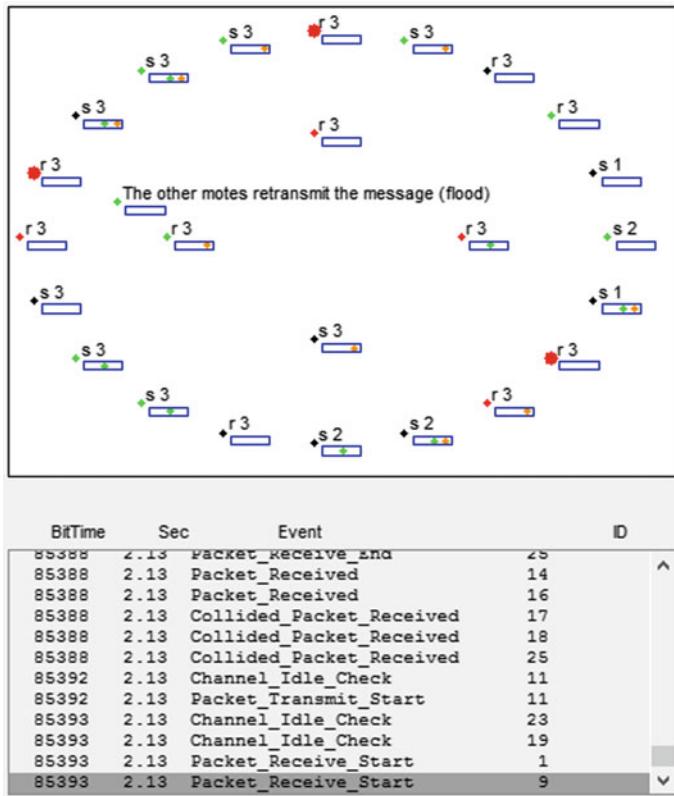


Fig. 1 User interface of PROWLER

system. WSNet is a discrete event-driven simulator which helps in implementing protocols. Both MATLAB and WSNet exchange information using sockets.

Most of the above-mentioned simulation environments have linear battery model. These models do not capture time-varying characteristics of the storage system that is dependent on energy harvesting. HarvWSNet's energy harvester model captures time-varying environmental parameters. Also, it has a power manager to define the charging and discharging characteristics of a battery. It also has provision of dual-power system. Depending on events and state of charge, network can switch between battery and photovoltaic. Other than above-mentioned simulation environments, there are many other simulators that can easily integrate with MATLAB [14].

4 Conclusion

This paper focuses on various simulation environments for wireless sensor networks. Beginning from basics, the paper also presents the extensions of each simulation framework. All the above simulation tools can be used for analysis of designed network before or after the deployment. Discussed simulation environments have different features to suit the different situations of various researchers. Table 1 summarizes and concludes the key features of above-discussed simulation environment.

Although latest simulation tools provide more features and compact protocols, lack of documentation limits its use. So, beginners can start with either NS-2 or QualNet. NS-2 has a wide range of the literature and examples available, but it demands good programming skills. QualNet is based on GUI, but it is a licensed simulator. Designers can also write their own simulation using MATLAB. Although MATLAB is a powerful tool, it does not support analysis of protocols at higher

Table 1 Comparison of simulation environments

Simulation environment	MAC protocol	Routing protocol	Energy consumption model	Radio model	Energy harvesting model
NS-2/Mannasim	IEEE 802.11, IEEE 802.15.4	DSDV, DSR, AODV, LEACH, directed diffusion	Yes	Free space and two-ray ground propagation	No
OMNeT++/Castalia	IEEE 802.11	INET support	Yes	Free space and two-ray ground propagation	No
GloMoSim/QualNet	IEEE 802.11, CSMA/CA	DSR, AODV	Yes	Two-ray ground propagation, log-normal shadowing, Longley-Rice, TIREM	No
TOSSIM	CSMA	Spanning tree	None	Probabilistic bit error	No
PROWLER/RMASE	IEEE 802.15.4	Ad hoc	Yes	Probabilistic	No
HarvWSNet	CSMA/CA, duty-cycling	Primitive (based on duty-cycling)	Yes	CC2500	Yes

levels of the protocol stack. Along with these simulation environments, testbeds like WISEBED [15] and Thorough IoT Testbed [16] can also be used for real-life experimentation.

References

1. Tonneau, A., Mitton, N., Vandaele, J.: How to choose an experimentation platform for wireless sensor networks? A survey on static and mobile wireless sensor network experimentation facilities. *Ad Hoc Netw.* **30**, 115–127 (2015)
2. Rekik, S., Baccour, N., Jmaiel, M., Drira, K.: Wireless sensor network based smart grid communications: challenges, protocol optimizations, and validation platforms. *Wirel. Pers. Commun.* Springer **95**, 4025–4047 (2017)
3. Issariyakul, T., Hossain, E.: Introduction to Network Simulator NS-2, 2nd edn. Springer, New York (2008)
4. Mannasim.: <http://www.mannasim.dcc.ufmg.br/>. Last accessed 30 July 2018
5. Park, S., Savvides, A., Srivastava, M.B.: SensorSim: A simulation framework for sensor networks. In: Proceedings of 3rd international ACM workshop on Modeling, analysis and simulation of wireless and Mobile systems, Boston, USA (2000)
6. Varga, A., Hornig, R.: An overview of the OMNeT++ Simulation Environment. In: Proceedings of the 1st international conference on simulation tools and techniques for communications, Networks and Systems & Workshops, vol. 60, pp. 1–10. France (2008)
7. Qualnet.: <http://web.scalable-networks.com/content/qualnet>. Last accessed 30 July 2018
8. Pediaditakis, D., Tsilishchev, Y., Boulis, A.: Performance and scalability evaluation of the castalia wireless sensor network simulator. In: Proceedings of 3rd international conference on simulation tools and techniques, Spain (2010)
9. Levis, P., Lee, N., Welsh, M., Culler, D.: TOSSIM : Accurate and scalable simulation of entire TinyOS applications. In: Proceedings of 1st international conference on embedded networked sensor systems, pp. 126–37. ACM, New York (2003)
10. Ali, Q.I.: Simulation framework of wireless sensor network(WSN) Using MATLAB/SIMULINK Software. MATLAB—A Fundamental Tool for Scientific Computing and Engineering Applications (2012)
11. ThingSpeak, <https://in.mathworks.com/help/thingspeak/>. Last accessed 30 July 2018
12. Simon, G., Volgyesi, P., Maróti, M., Lédeczi, Á.: Simulation-based optimization of communication protocols for large-scale wireless sensor networks. In: IEEE Aerospace Conference, vol. 3, pp. 1339–1346. USA (2003)
13. Didioui, A., Bernier, C., Morche, D., Senteiye, O.: HarvWSNet : A co-simulation framework for energy harvesting wireless sensor networks. In: international conference on computing, networking and communications, pp. 808–812. San Diego, CA, USA (2013)
14. Rajaram, M.L., Kougianos, E., Mohanty, S.P., Choppali, U.: Wireless sensor network simulation frameworks: A tutorial review: MATLAB/Simulink bests the rest. *IEEE Consum. Electron. Mag.* **5**, 63–69 (2016)
15. Chatzigiannakis, I., Fischer, S., Koninis, C., Mylonas, G., Pfisterer, D.: WISEBED : An open large-scale wireless sensor network Testbed. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 29, pp. 68–87. Springer, (2010)
16. Papadopoulos, G., Gallais, A., Schreiner, G., Jou, E., Noel, T.: Thorough IoT testbed characterization: from Proof-of-concept to Repeatable Experimentations. *Comput. Netw.* **119**, 86–101 (2017)

Design of an Architectural Framework for the Implementation of eHealth in Mauritius



Sudha Cheerkoot-Jalim and Kanishka Gopal

Abstract In Mauritius, people are becoming more and more health conscious, and as a result, their expectations from the health services are rising, thus compelling the government to re-think its health strategy. Taking into consideration the fact that the government's endeavour is to computerise the health services, this study proposes an architectural framework for the eHealth system in Mauritius. Different architectural approaches were studied, and an eHealth architecture for Mauritius has been proposed. A prototype of the patient management module was implemented to test and evaluate the proposed eHealth architecture. The prototype was thoroughly tested, and it has shown to be working smoothly on the proposed architecture. The result of this work can now be used by the Ministry of Health and Quality of Life as a foundation in embarking on the eHealth strategy of the government as part of its vision to computerise the health sector.

Keywords eHealth · mHealth · Electronic protected health information · Health level 7 · HIPAA · Electronic health records

1 Introduction

Information and communication technologies (ICT) have become an indispensable tool in peoples' lives, in the society and in the economy as a whole, and are being used in almost all spheres of life. Their adoption has now become a precondition for dynamic financial growth and future sustainability. At the government level, the objective today is to make maximum use of ICT in the delivery of services so as to achieve more efficiency and, thus meeting, to a better extent, the aspirations of the population.

S. Cheerkoot-Jalim (✉) · K. Gopal
University of Mauritius, Reduit, Mauritius
e-mail: s.cheerkoot@uom.ac.mu

K. Gopal
e-mail: kanishka.gopal1@umail.uom.ac.mu

In its e-Government strategy 2013–2017 [1], the Government of Mauritius has established policies for better adoption of innovative ICT solutions for the more efficient delivery of services to its citizens. At present, ICT is being successfully and extensively used in various sectors like education, training, employment and communication, and a number of services are now being offered through the electronic medium. The government believes that one of the key sectors in which there is much potential for applying ICT for more efficiency and for the betterment of the life of citizens is the health sector.

Across the globe, particularly in industrialised countries, the introduction of computer technologies in the health sector has brought many profound changes in healthcare. Authors of [2] have shown evidence that integration of ICT solutions in the health sector could considerably improve the quality and efficiency of healthcare services. In order to catch up with the rest of the world, the Government of Mauritius has come up with a project plan of computerising the health sector. The main objectives are to create a linkage and connectivity among various health institutions like hospitals and area health centres; providing a quality service to the public at large; bringing about more efficiency in healthcare; reducing waiting time for patients; keeping proper medical records; and reducing paperwork, amongst others [3]. This project is a very ambitious one which, however, is still at an embryonic stage.

1.1 Current Health Sector in Mauritius

The island of Mauritius had an estimated population of 1,221,150 residents at the end of 2016 [4]. The Ministry of Health and Quality of Life (MoHQL) is the main body responsible for the health status of the population. With a workforce of 1155 doctors, 3686 nurses and midwives, 69 dentists and 27 pharmacists in the public sector [4], free healthcare services are provided to Mauritian citizens in public health institutions, which include five regional hospitals, two district hospitals and five specialised hospitals. Two community hospitals, 5 mediclinics, 18 area health centres and 116 community health centres also deliver out-patient services. For the year 2016, there were 208,847 admissions and a total of 5.5 million cases seen by doctors in the out-patient service in the public sector [4].

With the considerable workload, and following numerous complaints from members of the public concerning service delivery in public health institutions [5], there is an ever-increasing demand for more efficient service delivery in the public health sector. The MoHQL has thus earmarked an eHealth project which is aimed at making use of cost-effective ICT solutions for better healthcare services, resource planning, interdepartmental and inter-sectoral collaboration. A preliminary design of the technology component in the eHealth plan [3] has been put up and is shown in Fig. 1. This diagram highlights the inputs and outputs, and the types of information systems and databases which the MoHQL had planned for the eHealth technical infrastructure. The project however is still at an embryonic stage and needs to be reignited.

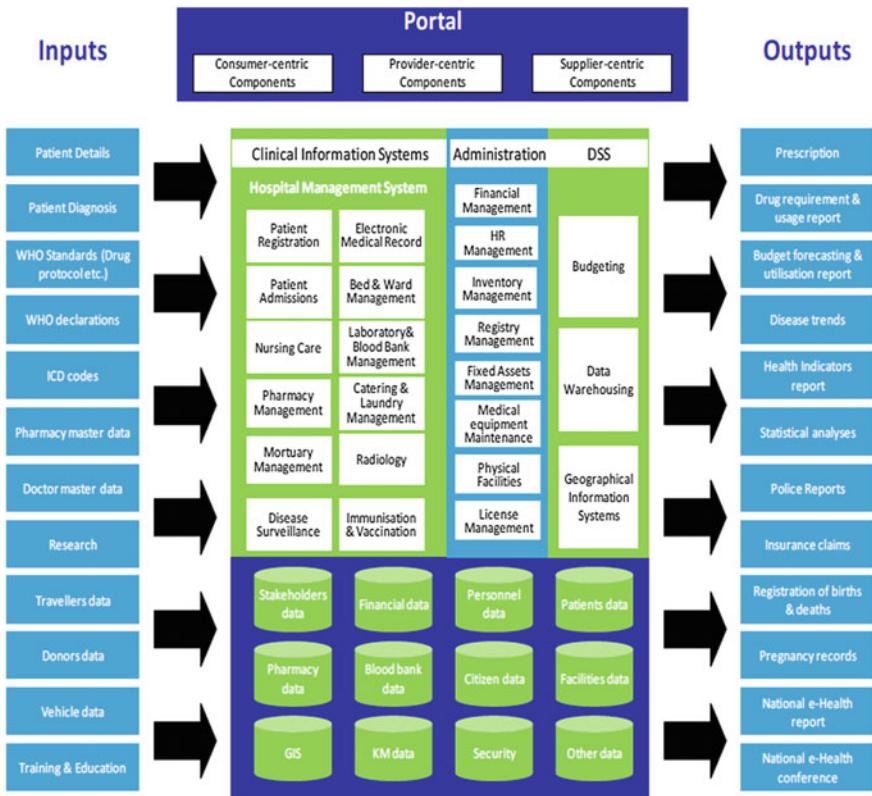


Fig. 1 An overview of the proposed technology infrastructure in the eHealth plan [3]

1.2 *Rationale of the Study*

The purpose and importance of software architecture are often highly underestimated. Software systems must be built on strong foundations due to their high level of complexity. For a project like the eHealth system which is of national interest, failing to design a suitable architecture at the initial stage can put the whole project at risk. Designing a suitable eHealth architecture will help minimise the risk of project failure and of having a poor software which is incapable of supporting future needs of the business environment. Therefore, in this study the different approaches used in the design of eHealth systems are studied. Based on the research work and the business and technical requirements for Mauritius, an architectural framework for eHealth has been proposed, considering crucial dimensions like security, confidentiality, interoperability and standards in healthcare.

2 Literature Review

The architecture of an ICT system includes important decisions related to the functional and non-functional requirements of the system, and these decisions can have a significant influence on the usability, performance, maintainability, security, economic and technology constraints and the overall success of the final product [6]. The aim behind designing an architecture for an application system is to mitigate the business risk associated with software development. The authors of [6] suggest that an architect should address the natural shift that will occur in the hardware and software technologies in the future when designing new software systems.

2.1 *eHealth Adoption*

eHealth strategies are being adopted in many countries around the world. Estonia and Canada are two examples of success stories in terms of the advancement in healthcare services due to the adoption of eHealth. The implementation of a national health information system by the Government of Estonia in the early 2000s was a means to streamline healthcare services and boost efficiency in healthcare procedures. It stands out first among the countries which implemented a successful eHealth system [7]. Since the eHealth plan was initiated in the year 2000, the three pillars have been the setting up of a “nationwide integrated electronic health record system”, the “standardisation” process with a focus on solving interoperability issues and the “integration of existing information technology infrastructure” using basic ICT platforms [7].

A study conducted by PricewaterhouseCoopers (PwC) on the adoption of electronic medical records in Canada [8] has found that \$1.3 billion were saved in six years that is between 2006 and 2012, through the implementation of community-based electronic health records. Moreover, the study has observed that the use of electronic medical records has remarkably enhanced the communication between healthcare providers and patients since medical information is provided to patients in real time. Through the Canada Health Infoway [8], the government is working on an integrated patient-centred eHealth system and a shared knowledge-based system for healthcare information so that meaningful data can be extracted from it to improve the treatment of diseases. The Canadian government has recognised that there is a must in expanding its digital health solutions for the federal population.

Some barriers are however identified for the implementation of eHealth in certain regions of the world, although its adoption is considered crucial. The authors of [9] have identified barriers which hinder the adoption of eHealth in the sub-Saharan region of Africa. These include the lack of proper funding for investment in eHealth and poor building infrastructure, the lack of familiarity with computers, no proper supply of electricity and regular power outages and government instability.

2.2 Related Work

In a study for improving the healthcare system through the use of ICT, the authors of [10] propose a three-tier cloud-based healthcare application architecture to be used to support an eHealth system. Their model is made up of a Rich Internet Application (RIA) on the client side for users to interact with the system through a cloud server. A logic layer is used to maintain the Web service, data query, application logic, transaction logic, data security, other functionality rules and third parties associated with the system. The three-tier architecture may allow important parts of medical data to be extracted from the cloud repository through data mining and be used to bring improvement in the healthcare system. However, this architecture has security challenges. In order to maintain data persistence, integrity, confidentiality and availability, a proper security mechanism has to be put in place. Moreover, the cloud service providers should agree on some open protocol and framework for secure information exchange.

The author of [11] presented a study on a national data centre solution and network architecture solution to host healthcare services and proposed the building of a national eHealth cloud environment to overcome many of the challenges confronting the implementation of the Hakeem eHealth project in Jordan. The author states that cloud computing creates new prospects for the eHealth system since any application hosted in the cloud can be easily updated to new standards, the speed of transferring medical information is improved, and it allows instant retrieval of medical information from different locations. Furthermore, services are provided at reduced costs. The system uses a centralised platform, since it allows the access to medical reports, body scans, electronic health records, prescriptions, patient data and medical history such as insurance claims, prescriptions and laboratory reports from any location. Moreover, the centralised data centre will reduce the risk of medical errors and wrong prescription of medicines and eliminate the problem of conflicting treatments with different doctors in different hospitals.

According to a study on cloud-based eHealth architecture for privacy-preserving data integration [12], it is argued that generalisation and pseudonymisation techniques can be used to ensure patients' data privacy on a cloud infrastructure. Health information can be independently and asynchronously exchanged, updated and searched in the cloud while ensuring the privacy of patients' information. Patients could make use of a smartcard to store a set of cryptographic keys for encryption/decryption of the sensitive data contained in their electronic health records (EHRs) in order to prevent unauthorised access. The proposed architecture in the study for privacy-preserving data integration consisted of a local database that would contain data about patients who receive treatment from a particular caregiver. A data repository is then hosted on an untrusted cloud server which stores electronic health records of patients, and anonymised patient data for the research purposes are stored on the cloud server on a research database.

3 Framework Development

An appropriate architectural framework for the implementation of eHealth has been designed and proposed for the Mauritian context and is detailed in the following subsections.

3.1 Methodology

The Centre for eHealth Research and Disease Management (CeHRes) roadmap [13] has been applied in this study to design the eHealth architecture. The CeHRes roadmap is a method used to plan, coordinate and execute the development process of eHealth by conducting holistic research in electronic health field and by the development of an eHealth technology that meets user requirements. It consists of five steps, and formative evaluation is performed at each step of the process in order to improve and produce the right eHealth technology. The five steps as applied to this work are listed below.

Contextual Inquiry. A literature review was conducted to obtain a thorough understanding of eHealth and architectures related to eHealth. Research was also conducted on the current ICT state in the healthcare sector in Mauritius.

Value Specification. Formal and informal meetings were held with various stakeholders at the MoHQL and with patients and doctors in different hospitals. Visits were also conducted in two major public hospitals in Mauritius to know more about hospital procedures and to check the feasibility of the eHealth project in Mauritius. The views of patients, doctors and other hospital staff on the use of ICT tools in hospitals were acquired. Problems being faced by the current system were also gathered during the interview process.

Design. Based on the literature review conducted and requirements obtained, an architectural framework for eHealth has been designed. This architecture required reworks and refinement as several design dimensions such as the cost of applying the architecture, number of users involved, security considerations and availability of information had to be taken into consideration. Standards and design quality criteria also had to be considered.

Operationalisation. A patient management module of the eHealth system has been implemented as a prototype to test and evaluate the proposed architecture. It consists of the management of patients, doctors, diagnosis, treatments and appointments.

Summative Evaluation. The prototype has been built on top of the architecture and evaluated for all the requirements of the eHealth system, in terms of usability, efficiency, security, interoperability and scalability.

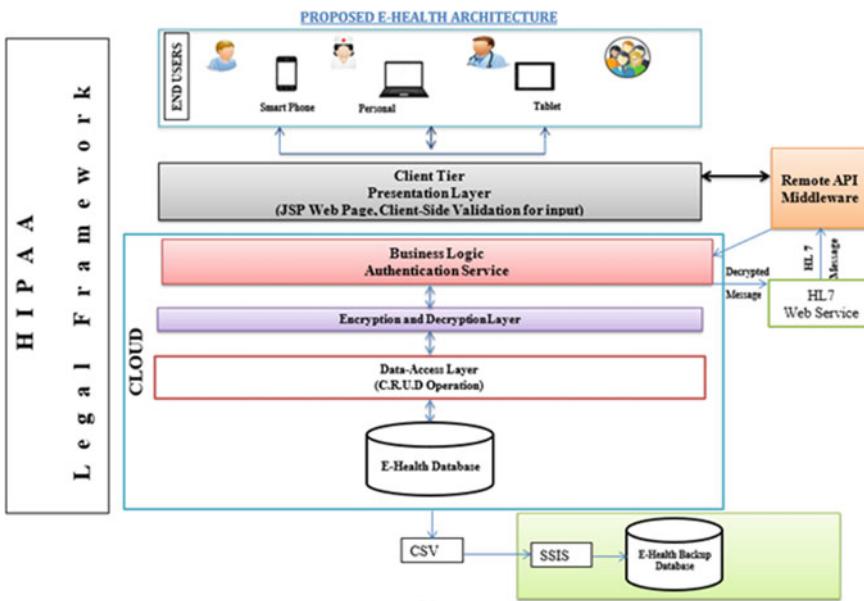


Fig. 2 Proposed eHealth architecture

3.2 Proposed eHealth Architecture

After a thorough requirement analysis for eHealth for Mauritius, an appropriate eHealth architecture has been designed, taking into consideration aspects like security, confidentiality, interoperability and standards in healthcare. The proposed architecture is shown in Fig. 2.

The different components of the architecture are described below.

Layers. The proposed architecture is a three-tier architecture with the following layers.

Presentation layer. This layer holds the different Web pages through which a user may interact with the system.

Business Logic Layer. This layer prescribes how components interact with each other, controls how and what data processing should be performed and determines the routes through which information should flow.

Data Access Layer. This layer is made up of different classes and procedures to perform C.R.U.D (Create, Read, Update and Delete) transactions.

Cloud-Based Storage Infrastructure. The architecture uses Google App Engine cloud datastore to keep all electronic protected health information.

HIPAA Security Rule. The proposed eHealth architecture comprises Health Insurance Portability and Accountability Act (HIPAA) security rule technical safeguards to protect the confidentiality of medical data on the application. This ensures that only authorised users can access the medical information. Theft of medical data

is also prevented, and any possibility of data loss is limited. The technical safeguards that the architecture is made up of are namely: access control, audit controls, integrity control, transmission security and encryption mechanisms.

HL7 Version 2 Standard. The system uses HL7 (Health Level Seven International Standard) version 2 messaging techniques to manipulate data pertaining to the patient. This technique involves retrieving patient data from the database, converting them into HL7 v2 messages by placing the required data in a specific order and separating them with pipelines. The message is then placed into a hash map data structure which, in turn, allows the data to be retrieved by using a specific key.

Middleware. To allow different Web applications to share health information, remote API is used as middleware. The different health institutions that have to share health data need to have the remote API plugged into their system. Additionally, a p12 file is generated by Google cloud platform and contains an encrypted secret key to allow the exchange of information between the different institutions in JSON format.

mHealth. With the remote API middleware plugged into the application and the p12 file set up to handle authentication and exchange of information, a mobile application can easily access data from the architecture's data store. Since the application is developed on cloud using platform-as-a-service, a view is created into an Android screen, which allows it to access the cloud application.

SSIS feature. SQL Server Integration Service (SSIS) is used to create a backup of the cloud system. The cloud system has the ability to export .csv files which are fed into the SSIS package. The SSIS package reads all .csv files from a folder location, identifies the files by their name and then transfers the data into the SQL server database.

4 Evaluation

One module of the eHealth system, the patient management module, was implemented and used as a prototype to test and evaluate the proposed architectural framework. The patient management module consists of managing medical information about patients, doctors, appointment details, diagnosis, observations, complaints, amongst others. The proposed architecture was evaluated according to the following criteria.

Centralised Database. The proposed architecture solves the issue of manually transferring files from one healthcare institution to another since all medical information will be stored electronically in a centralised database. By proposing Google App Engine which is a cloud application to store eHealth information, availability of eHealth data is guaranteed. Even when a patient changes doctor or changes hospital, his/her medical information can be retrieved. Furthermore, all information is stored in a more secured manner and ensures that health data is not intercepted while being transferred from one server to another by using “https” protocol. This architecture facilitates the transfer of data between different health centres.

Interoperability. The proposed eHealth architecture uses a middleware API and HL7 V2 standards. This shows that medical information can be accessed from any device (a personal computer, a laptop or a smartphone) from anywhere and at any time, but with the proper access rights. With the proposed architecture, patients do not have to move to hospitals and queue up for several hours only to check their appointment dates. Health Level 7 standards make the eHealth architecture compatible with other electronic health applications. For example, another health information system such as the blood laboratory system can query the eHealth system to check a patient's medical history within seconds without requiring hospital staff to manually look for patients' files. This improves healthcare delivery by providing patients with a more efficient service.

Customisation. The proposed architecture can be used in any hospital and can be adapted to a hospital's needs. An administrator will have the ability to control how the interface will look like in terms of colours and home page display. The administrator can also customise which users have access to which features.

Flexibility. The eHealth architecture dynamically generates the contents of the web pages based on values returned from the cloud data stores. For example, the navigation menu is populated based on the access level and pages defined in the user interface table. Thus, in the future, a new page and the option can be added to the architecture by an administrator himself and this can be done without an IT engineer's intervention. Additionally, the application, business and data access layers are loosely coupled, which makes the modification and addition of modules easier. For instance, the architecture can easily be modified to use cloud SQL instead of cloud datastore. The eHealth architecture can be adapted to emerging technologies in the future. One such example is the creation of an external database which can backup and archive data using SSIS for medical research. New technologies such as "big data" can be added to the architecture without disrupting the architectural flow of information.

Availability and Scalability. The proposed eHealth architecture makes use of Google App Engine (GAE) as cloud service provider. GAE has been configured with a cross-region load balancer. This will redirect traffic to another VM machine in another region set up on the cloud in the event of too much traffic on one server or unavailability of servers in one region. This guarantees the availability of health data in the event of an emergency. Furthermore, since the eHealth project is a complex one, consisting of several modules, the architecture has been designed with an application layer. This layer allows several more modules to be added to the architecture to support a fully integrated eHealth system in the future.

Security. The proposed architecture fully complies with HIPAA Security Rule Technical Safeguards. The five standards listed under the HIPAA Security Rule technical safeguards [14] have been fully met by the architecture as follows:

- **Access Control.** A unique username is assigned to each user based on his/her access level and the system controls which activities each user can perform based on the assigned access level.

- *Audit Control.* The architecture keeps track of the user's attempt to log into the system. This involves storing the username, IP address, login status, date and time.
- *Integrity.* Each time a user modifies a health-related data, the existing record and a flag is stored into a transaction audit table.
- *Authentication.* A username and password are mandatory to be able to access the architecture.
- *Transmission Security.* The architecture has been set up to transfer data using "https" protocol so that data is not intercepted during transmission.

mHealth. The architecture caters for mobile applications, and thus, end users have the ability to access health data at their fingertips. For instance, patients can view their appointment details and their medical history using their mobile phones. A doctor can also access patients' data from mobile phones in the event of emergency.

Migration of Health Data and Backup. The architecture allows the export of cloud health data in the format of flat files. The data can then be imported into another data source by the use of SQL Server integration packages. The SSIS package detects the name of the flat file and determines in which table/s to migrate the data to the local database. This will resolve issues such as transferring large amounts of data from one cloud platform to another, or to archive historical patient data into a local database which can later be used for research purposes.

5 Conclusion and Recommendations

As already elaborated in this study, the trend worldwide in the health sector is the adoption of eHealth infrastructures and systems to enable safer, more efficient, high-quality and customised healthcare services to be provided to citizens. Nowadays, the population is very demanding and health conscious and wants to see changes that would impact favourably on quality of life. Mauritius has embarked on an eHealth strategy, albeit progress is at a very slow pace. The positive point is that the Mauritian population is IT literate and adapts rather easily to changing technologies. This is depicted by the fact that the need for online application for an appointment at the hospital has been strongly expressed by the citizens [3].

In this study, an architectural framework for the implementation of an eHealth system for Mauritius has been proposed, taking into consideration the requirements of the Mauritian health sector and criteria like security, availability, confidentiality, interoperability and scalability. One of the modules of the eHealth system, which is the patient management System, has been developed and seamlessly mounted on the proposed eHealth architecture. After rigorous testing, it has shown to meet all requirements identified at the start of the study. The functionalities under this module, which include the registration of patients, input of patients' personal and medical information, scheduling of appointments, retrieval of information from anywhere in real time, among others, are working smoothly on the architecture and, with the appropriate infrastructure, can be successfully deployed.

The proposed architecture will be a good starting point for the eHealth strategy for the Ministry of Health and Quality of Life and the country as a whole in view of its specificities, like the use of open source technologies such as Google App Engine; adaptability to future technologies such as big data; instant access to health information, given that smartphones are extensively being used by the Mauritian population. From an administrative perspective, there will be drastic reduction in costs, less wastage of resources, less hassle for patients, better management at the level of the health institutions. Moreover, management will be better equipped to take appropriate policy decisions regarding the health sector. As a whole, the architecture developed will be a solution to the problem areas identified in the present health set up and which the Ministry is trying to resolve in order to provide a better, efficient and personalised health service to the citizens.

Recommendations. For any new system to be implemented successfully and to enable it to generate the expected efficiencies, it should be accompanied by some measures to support it, and this is applicable to the eHealth project as well, which will be a major transformation in the health sector. A few recommendations include:

- Investment in the appropriate ICT infrastructure required to technically support the proposed architecture;
- Awareness campaigns to encourage the acceptance of the system by the health sector staff and citizens;
- Appropriate training of health sector staff;
- Assistance to vulnerable groups for the use of the system;
- Continued evaluation and feedback on the system.

The timely implementation of these recommendations will ensure that the system is duly accepted and used, and generates the expected benefits. Though the Ministry of Health and Quality of Life has a long way to go regarding the eHealth project, the political will is there. So, there is no doubt that with the required resources and the collaboration of all stakeholders, the project will be a reality in the near future, and eHealth will go a long way in improving the efficiency of healthcare in Mauritius and, eventually, promoting good health of the population.

References

1. Central Informatics Bureau.: Ministry of Information and Communication Technology: E-Government Strategy 2013–2017 (2013)
2. Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., Shekelle, P.G.: Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Intern. Med.* **144**(10), 742–752 (2006)
3. Ministry of Health and Quality of Life.: Health 2015: Seamless Continuity of Care (2010)
4. Ministry of Health and Quality of Life.: Health Statistics Report 2016. (2017)
5. Bholah, L.A., Beharee, K.: Mauritius eHealth-Trust in the healthcare revolution. In ITU Kaleidoscope: Trust in the Information Society (K-2015) pp. 1–8. IEEE (2015)
6. Shaw, M., Garlan, D.: Software Architecture: Perspectives on an Emerging Discipline. Prentice Hall, Englewood Cliffs (1996)

7. Sabes-Figuera, R., Maghiros, I.: European Hospital Survey: Benchmarking Deployment of e-Health Services 2012–2013. European Comission (2013)
8. PricewaterhouseCoopers.: Canada Health Infoway: The emerging benefits of electronic medical record use in community-based care (2013)
9. Odekunle, F.F., Odekunle, R.O., Shankar, S.: Why sub-Saharan Africa lags in electronic health record adoption and possible strategies to increase its adoption in this region. *Int. J. Health Sci.* **11**(4), 59 (2017)
10. Biswas, S., Akhter, T., Kaiser, M.S., Mamun, S.A.: Cloud based healthcare application architecture and electronic medical record mining: an integrated approach to improve healthcare system. In: 17th international conference on computer and information technology (ICCIT), 2014, pp. 286–291. IEEE (2014)
11. Alzoubaidi, A.R.: Cloud computing national e-health services: data center solution architecture. *Int. J. Comput. Sci. Network Secur. (IJCSNS)* **16**(9), 1 (2016)
12. Dubovitskaya, A., Urovi, V., Vasirani, M., Aberer, K., Schumacher, M.I.: A cloud-based eHealth architecture for privacy preserving data integration. In: IFIP international information security conference, pp. 585–598. Springer, Cham (2015)
13. Van Gemert-Pijnen, J.E.W.C.: The CeHRes roadmap. In: Van Gemert-Pijnen, JEWc, Peters, O., Ossebaard, HC.(eds.) Improving Ehealth, vol. 9, pp. 33–48. ISBN: 978-94-6236-067, (2013)
14. Centers for Medicare & Medicaid Services, Department of Health and Human Services: Security standards: technical safeguards. HIPAA Security Guidance for Remote Use of and Access to Electronic Protected Health Information, 2 (2007)

Sentiment Analysis on Online Product Reviews



Rajesh Bose, Raktim Kumar Dey, Sandip Roy and Debabrata Sarddar

Abstract Today, people are exchanging their thoughts through online Web forums, blogs, and different social media platforms. Sometimes, they are giving reviews and opinions on different products, brand, and their services. Their reviews toward a product not only improve the product quality but also influence purchase decisions of the consumers. Thus, product review analysis is a widely accepted platform where consumer can easily aware of their requirements. In this experiment, we track 568,454 fine food reviews of 74,258 products and 256,059 users on Amazon over a period of ten years. To analyze the result, we select six most popular products and users based on the plain text review, and NRC emotion lexicon is used which can be categorized eight basic emotions and two sentiments. Word cloud also help our research to make comparisons between the eight emotion categories. Our results show that how sentiment analysis will help to identify the consumers' behaviors and overcome those risks to meet the consumers' satisfaction.

Keywords Amazon's customer reviews · Digital marketing · Electronic word-of-mouth (e-WOM) · Sentiment analysis · Unstructured data

R. Bose · R. K. Dey
Simplex Infrastructures Ltd, Kolkata, India
e-mail: bose.raj00028@gmail.com

R. K. Dey
e-mail: deyraktim@gmail.com

S. Roy (✉)
Brainware Group of Institutions, Kolkata, India
e-mail: sandiproy86@gmail.com

D. Sarddar
University of Kalyani, Kalyani, India
e-mail: dsarddar1@gmail.com

1 Introduction

Today, electronic word-of-mouth (e-WOM) is one of the most important factors for digital marketing [1]. Companies are using the digital platforms for promoting their products. Nowadays, customer's online reviews can influence the purchase decisions of a product. Different e-commerce companies like Amazon, eBay, Flipkart, and Walmart are analyzing their customers' reviews in different ways [2, 3].

Unstructured data such as text, audio, photos, or videos are posted by the customers in social media Web sites, Web forums, blogs, and different online review platforms [4, 5]. Natural language processing (NLP) is an automatic programming technique to analyze and comprehend large amounts of customers' opinions where different companies build chatbots to support online customer service interactions [6–8].

Sentiment classifications play an important role in classifying unstructured data. Cui et al. classified online product reviews into positive and negative classes [9]. In their experiment, they used different machine learning algorithms to evaluate different trade-offs of 100 K of online product reviews. Fang et al. tackled the sentiment polarity categorization problem in online product reviews of Amazon.com [10]. They categorized their outcomes into two levels: (a) sentence-level categorization and (b) review-level categorization.

In this manuscript, we used NRC emotion lexicon that categorized customers' reviews into eight emotions (anger, fear, trust, anticipation, sadness, surprise, disgust, and joy) and two sentiments (positive and negative) [11]. Our proposed analyzing system can enhance the review technique, which has been discussed by Escalona [12].

The detailed sections of the manuscript are as follows. Section 2 describes the literature survey. Section 3 deals with data preparation along with review analysis. Result analysis part is discussed in Sects. 4 and 5 and draws conclusions and examines the possibilities of next future.

2 Literature Survey

Information can be classified into two basic statements: (a) objective statements and (b) subjective statements. Subjective statements are the opinions that consider people's sentiments about entities and events [13]. Bhatt et al. proposed a system that classified the customer reviews along with the sentiment of the reviews [13].

Patel proposed a machine learning technique that classifies sentiments of review based on words, i.e., positive or negative [14]. Sykounas et al. tested his work on the SemEval Twitter-based corpora and ESWC semantic sentiment analysis challenge which get the result of 88.05% accuracy over the test set for detecting the positive and negative sentiment [15].

In [16], authors have shown that support vector machine (SVM) performs better accuracy compared to Naïve Bayes and maximum entropy methods. Leung et al. also classified the sentiment into two classes as positive and negative [17].

Ray et al. proposed a framework for sentiment analysis using R software. In their method, they collected the plain text from Twitter, pre-processed, and then lexicon-based approach is used for analyzing user's sentiment [18].

Govindaraj [19] proposed a method which extracts two sets of features from the customer reviews: acoustic features which represent emotions and lexical features that represent sentiments. These sets are used in a supervised classifier to predict the sentiment of the customers. In their experiments, they used an audio speech dataset prepared and downloaded from Amazon and YouTube online portal, respectively. In the experiment [20], the author designed a text classifier using Python Data Analysis Library, Natural Language Toolkit, and Scikit-learn libraries. McAuley et al. tested that open-ended queries performed better than binary queries [21].

Babu proposed a sentiment analysis technique for e-shopping Web sites [22]. He used SentiWordNet dictionary to find scores of each word [23, 24]. Keith et al. presented a combined approach of unsupervised machine learning along with the techniques from NLP [25]. The authors also have a comparison study with classical machine learning methods such as support vector machine (SVM) and Bayesian classifier (NB). Li also proposed a framework to identify two stages: feature extraction and classifier learning [26, 27]. Bose et al. proposed the emotion analysis using syuzhet in CRAN package which is based on NRC emotion lexicon [28]. In our proposed method, we used the same package to categorize emotion into eight sub-categories and two sentiments [28].

3 Data Preparation

Dataset has collected from the Kaggle which consists of reviews of fine foods from Amazon from October 1999 to October 2012 [29, 30]. Dataset includes 568,454 reviews, 74,258 products, 256,059 users, and 260 users with more than 50 reviews.

4 Data Pre-processing

Dataset in the form of plain text review is mainly unstructured data [31]. At first, we apply data pre-processing method on our dataset [32, 33]. In this data pre-processing steps, we have followed the following steps [34]:

- Remove all URLs (e.g., www.abc.com), screen name (e.g., @username), and all hashtags (e.g., #topic)
- Remove all symbols, punctuations, and numbers
- Remove stopwords

- Substitute any non UTF-8 by space
- Replace all the emoticons with their sentiment
- Change text to lowercase
- Replace words with their stems or roots
- Remove the retweets.

5 Product Review Analysis

In our experiment, NRC emotion lexicon, large word list (like other lexicon AFINN, ANEW, EmoLex, LabMT, General Inquirer, and SentiWordNet), is used for sentiment analysis [11, 35]. This lexicon (version 0.92) had about 14,200 unigrams word types with word-level emotion [11, 36].

In Figs. 1 and 2, bar charts represent the eight emotions (anger, fear, trust, anticipation, sadness, surprise, disgust, and joy) and two sentiments (positive and negative). We track six most popular users and products for our sentiment analysis.

For user-centric approach, we found a potential reviewer who made their own opinion to the m number of products on the Amazon. Our aim is to find out the sentiment based on their reviews. These analyses will help to identify the sentiment of the customers. Similarly, in product-centric approach, we found the best-reviewed product which has been reviewed by n number of customers. We also analyzed the sentiment or emotion of the n number of customers toward the products.

In Figs. 3 and 4, word cloud is used to visualize review data which illustrates the keywords used more frequently by the customers. We categorized our reviewed text into eight emotion categories [37, 38]. In each category, highest-frequency word will identify the most insightful of the customers [39, 40]. In our experiment, most of the Amazon customers reviewed by the below keywords: e.g., “good” or “wonderful” in surprise section and “flavor” or “organic” or “smell” in disgust section.

6 Result Analysis

In this manuscript, we analyzed the sentiment or emotion of the products based on customers reviews of Amazon.com. Our aim of this research is to identify the highest reviewed product and also track the most anxious word in each emotion category of the customers using the word cloud.

In the Figs. 1 and 2, most of the popular product reviews of Amazon users and Amazon products are shown, respectively. Most of the reviews have more polarity for “positive,” “joy,” and “trust” rather than “fear,” “anger,” and “disgust.”

Differently, Fig. 5 shows that “negative,” “disgust,” and “anger” have more polarity than “trust,” “joy,” and “surprise” for the sentiment analysis of the least popular product reviews in Amazon.com.

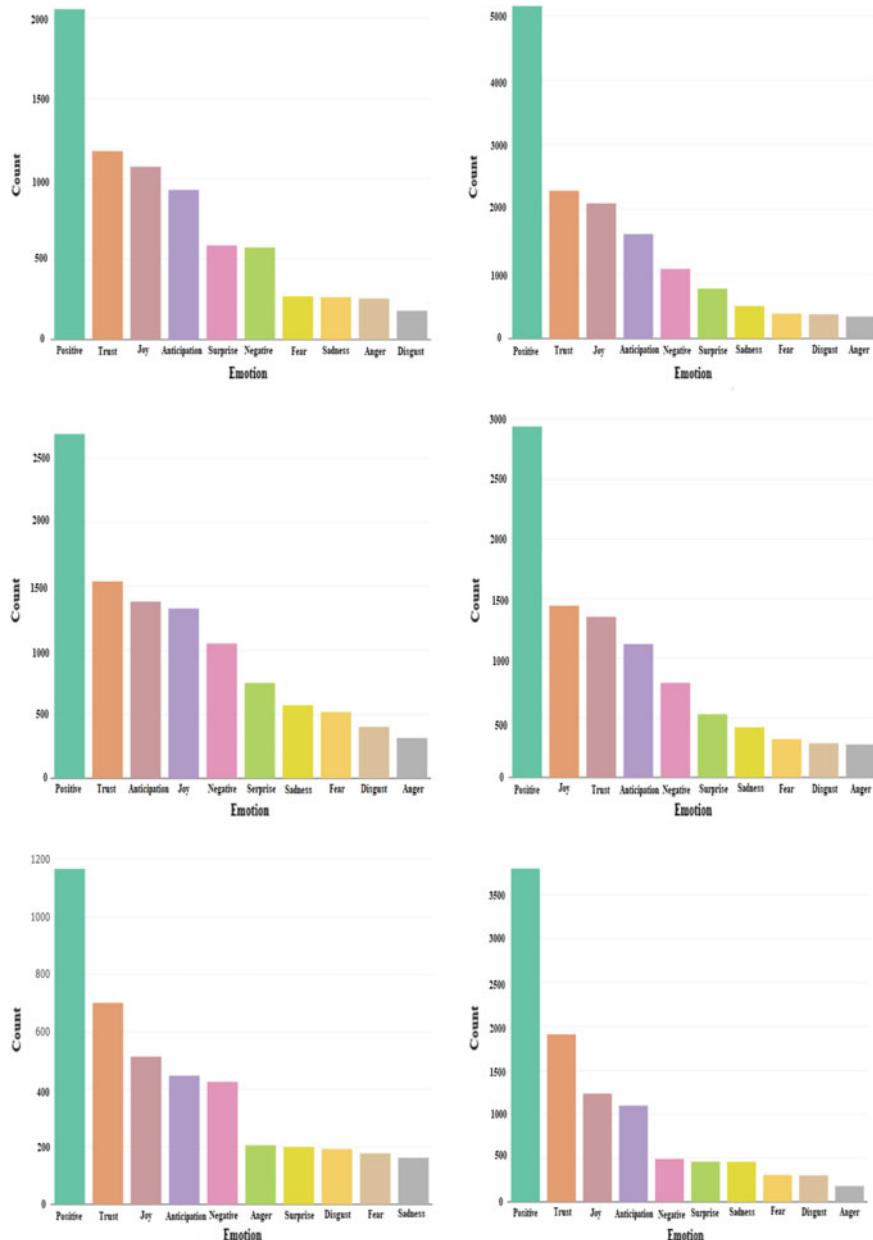


Fig. 1 Distribution of emotion of six most popular product reviews of Amazon users

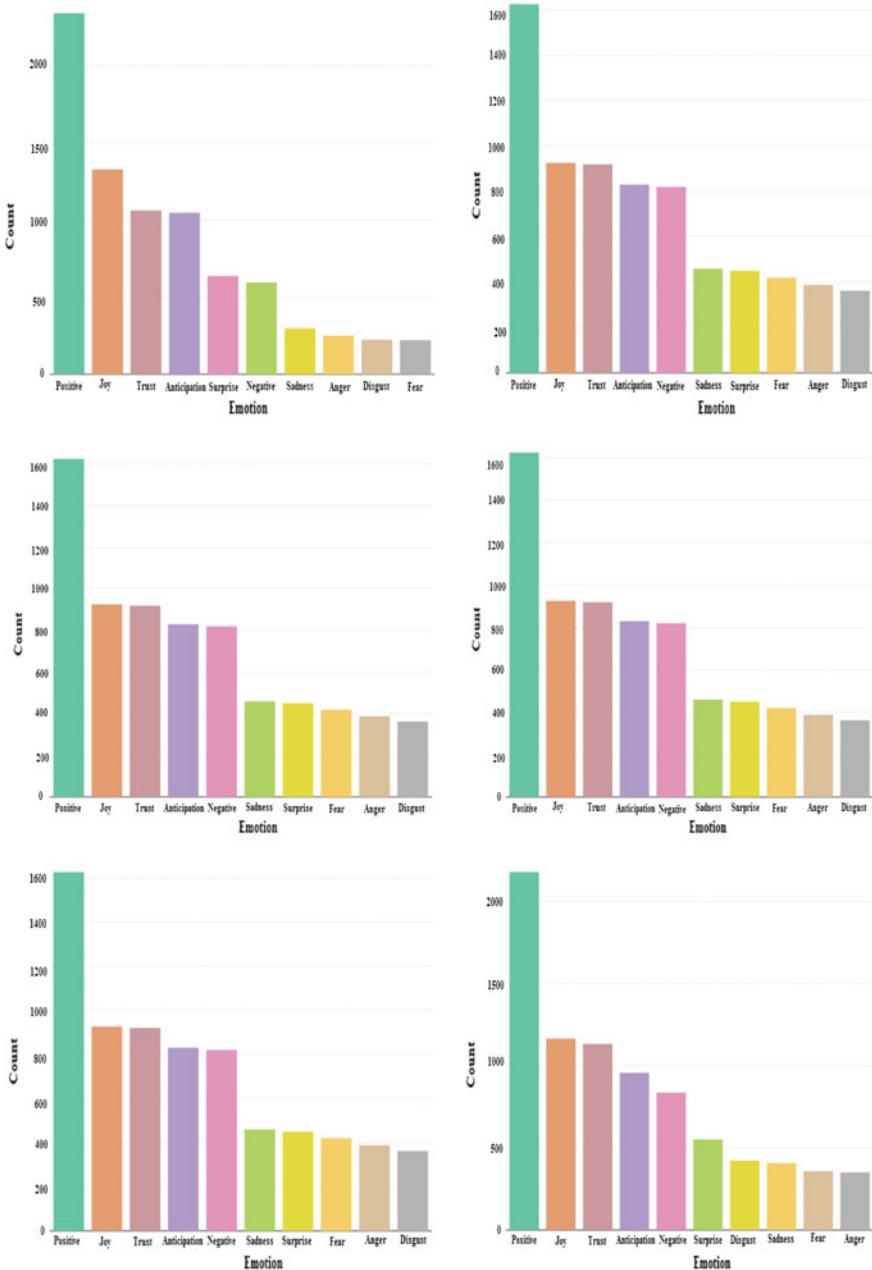


Fig. 2 Distribution of emotion of six most popular product reviews of Amazon products

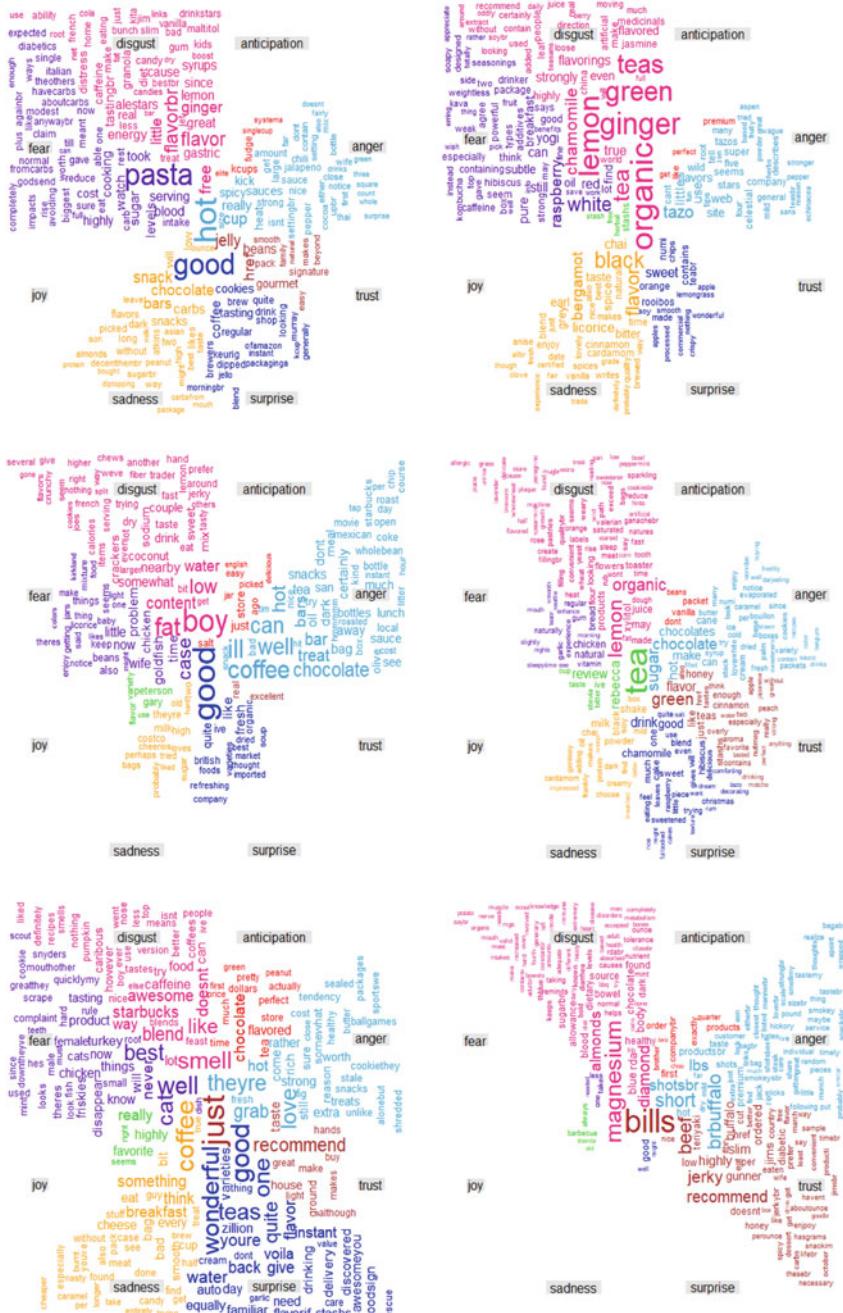


Fig. 3 Word Cloud of six most popular product reviews of Amazon users

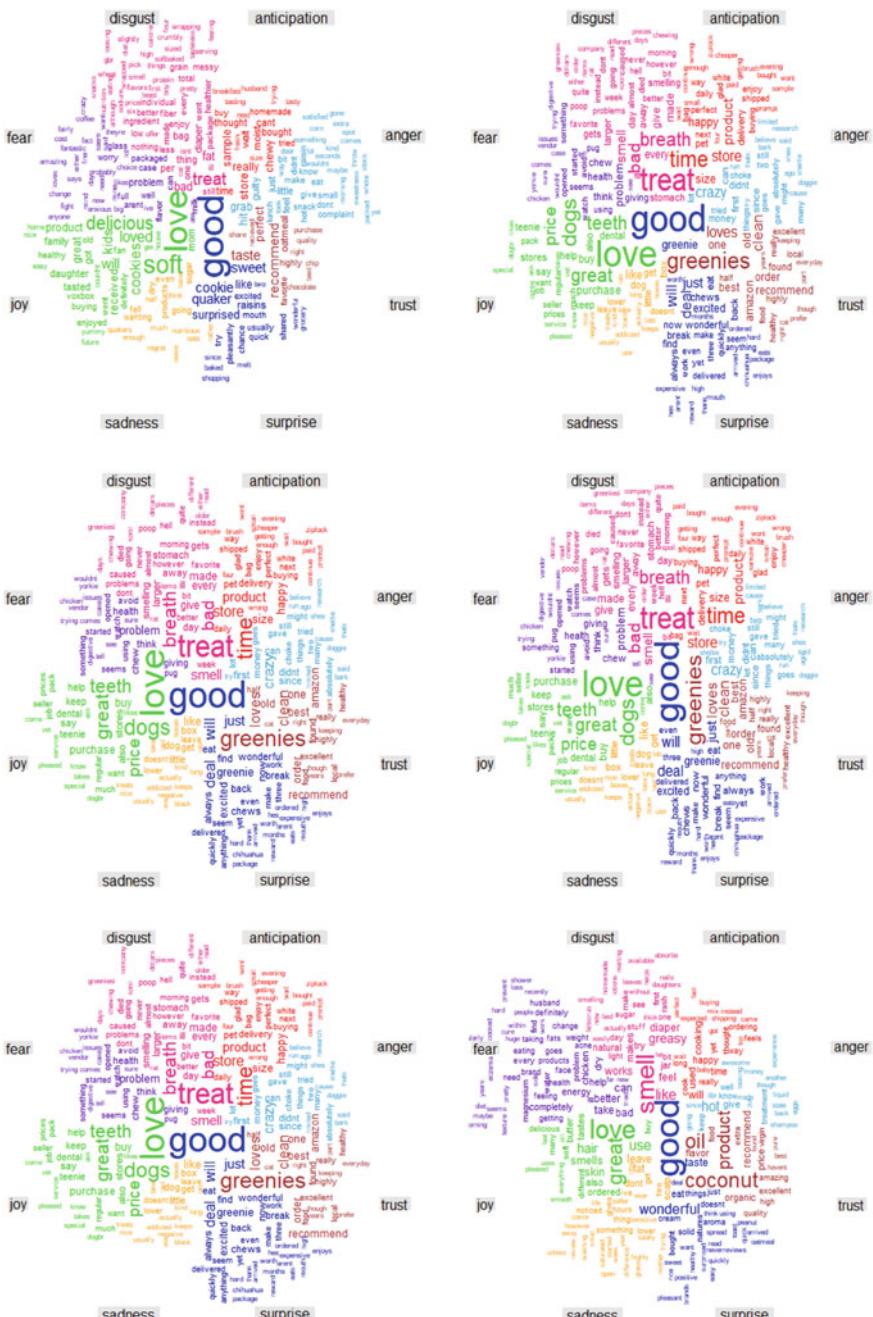


Fig. 4 Word Cloud of six most popular product reviews of Amazon products

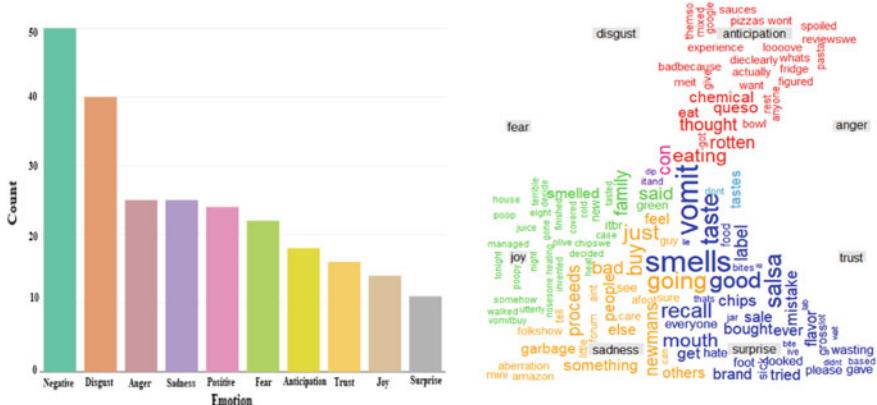


Fig. 5 Sentiment analysis of least popular product reviews of Amazon product

Furthermore, authority of the Amazon product has more concern about those words, e.g., “vomit,” “smells,” and “rotten” that users are more conscious about the quality of the product.

7 Conclusion and Future Work

In our research, we have done the sentiment or emotion classification using NRC emotion lexicon. Our experiment results show that how customers’ reviews are important for digital marketing research. Sometimes, customers’ given ratings differ with their comment. Researchers work on the sentiment analysis of the customers’ reviews of different e-commerce Web sites.

Our research is not free from constraints. Future work may be performed as below: We will include some topic modeling-based sentiment analysis features on the plain text reviews of the customers on e-commerce Web sites, which will predict the best product as per the customers’ needs. We will also perform perplexity analysis and lexicon quality assessment on plain text review data of the customers in the near future.

References

1. Wang, W., Wang, H., Song, Y.: Ranking product aspects through sentiment analysis of online reviews. *J. Exp. Theo. AI.* **29**, 227–246 (2017)
2. Waldron, M.: Building a text analysis process for customer reviews in RapidMiner. AYLIEN (2016)
3. Spiewanowski, P.: How to Visualize and Analyze Customer Feedback. uxBOOTH (2016)
4. Aswani, S.: Analyzing Customer Feedback Data: Manual Analysis vs NLP. CLARABRIDGE (2017)
5. Luminoso.: Customer Feedback and Artificial Intelligence. White Paper (2016)
6. Orf, D.: Google assistant is a mega AI Bot that wants to be absolutely everywhere. Google (2016)
7. Thematic.: Visualizing Customer Feedback: Three Alternatives to Word Clouds. Medium (2017)
8. Abuleil, S., Alsamara, K.: Using NLP Approach for analyzing Customer Reviews. SOEN—2017, 117 – 124 (2017)
9. Cui, H., Mittal, V., Datar, M.: Comparative Experiments on Sentiment Classification for Online Product Reviews. American Association for AI, pp. 1265–1270 (2006)
10. Fang, X., Zhan, J.: Sentiment analysis using product review data. *J. Big Data.* **2**(5), 1–14 (2015)
11. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word-emotion association Lexicon. *Comput. Intell.* **29**(3), 1–25 (2013)
12. Escalona, T.: Detect sentiment from customer reviews using Amazon Comprehend. Amazon Machine Learning Blog (2018)
13. Bhatt, A., Patel, A., Chheda, H., Gawande, K.: Amazon review classification and sentiment analysis. *IJCSIT* **6**, 5107–5110 (2015)
14. Patel, B.: Predicting Amazon product reviews' ratings. Medium (2017)
15. Sygkounas, E., Rizzo, G., Troncy, R.: Sentiment polarity detection from amazon reviews: an experimental study. SemWebEval 2016. *Commun. Compt. Info. Sci.* **641**, 108–120 (2016)
16. Shivaprasad, T.K., Shetty, J.: Sentiment analysis of product reviews: a review. ICICCT, India (2017)
17. Leung, C.W.K., Chan, S.C.F.: Sentiment analysis of product reviews. *Encycl. Data Warehous. Mining.* 1794–1799 (2009)
18. Ray, P., Chakrabarti, A.: Twitter Sentiment Analysis for Product Review Using Lexicon Method. ICDMAI, India (2017)
19. Govindaraj, S., Gopalakrishnan, K.: Intensified sentiment analysis of customer product reviews using acoustic and textual features. *ETRI J.* **38**, 494–501 (2016)
20. Rajkumar, V.: Sentiment analysis for Yelp review classification. Medium (2017)
21. McAuley, J., Yang, A.: Addressing complex and subjective product-related queries with customer reviews. WWW, Canada (2016)
22. Babu, U.R.: Sentiment analysis of reviews for e-shopping websites. *IJECS.* **6**, 19965–19968 (2017)
23. Esuli, A., Sebastiani, F.: SENTIWORDNET: A publicly available lexical resource for opinion mining. In: Proceedings of the 5th conference on language resources and evaluation, pp. 1–6 (2016)
24. Baccianella, S., Esuli, A., Sebastiani, F.: SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: Proceedings of the 5th conference on language resources and evaluation, pp. 1–5 (2010)
25. Keith, B., Fuentes, E., Meneses, C.: A hybrid approach for sentiment analysis applied to paper reviews. In: Proceedings of ACM SIGKDD Conference, pp. 1–10. Halifax, Nova Scotia, Canada (2016)
26. Neo4 J.: Sentiment Analysis for Amazon Reviews using Neo4 J. Kvangundy (2015)
27. Li, L.: Identification of informative reviews enhanced by dependency parsing and sentiment analysis. ICCCI, China (2016)

28. Bose, R., Dey, R.K., Roy, S., Sarddar, D.: Analyzing Political Sentiment using Twitter Data. ICTIS, India (2018)
29. McAuley, J., Leskovec, J.: From Amateurs to Connoisseurs: Modeling the Evolution of User Expertise Through Online Reviews. WWW, Canada (2013)
30. SNAP:: Amazon Fine Food Reviews. <https://www.kaggle.com/snap/amazon-fine-food-reviews>. Kaggle (2018)
31. Geetha, M., Singha, P., Sinha, S.: Relationship between customer sentiment and online customer ratings for hotels—An empirical analysis. *Tourism Mgmt.* **61**, 43–54 (2017)
32. Krouská, A., Troussas, C., Virvou, M.: The effect of preprocessing techniques on Twitter sentiment. IISA, India (2016)
33. Zin, M.H., Mustapha, N., Murad, M.A.A., Sharef, N.M.: The effects of pre-processing strategies in sentiment analysis of online movie reviews. In: AIP conference proceedings 1891, 020089 (2017)
34. Hemalatha, I., Varma, G.P.S., Govardhan, A.: Preprocessing the informal text for efficient sentiment analysis. *IJETTCS* **1**, 58–61 (2012)
35. Word list. http://neuro.imm.dtu.dk/wiki/Word_list (2018)
36. Mohammad, S., Turney, P.: Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon, pp. 1–9. NAACL-HLT, LA, California (2010)
37. Nguyen, T.T., Chang, K., Hui, S.C.: Word Cloud Model for Text Categorization. ICDM, Canada (2011)
38. Vithiya Ruba, K., Venkatesan, D.: Building a customer sentiment analysis tool based on an ontology for Twitter posts. *IJUST* **8**, 1–9 (2015)
39. Rajput, Q., Haider, S., Ghani, S.: Lexicon-based sentiment analysis of teachers' evaluation. *Appl. Comput. Intell. Soft Comput.* **2385429**, 1–12 (2016)
40. Rambocas, M., Gama, J.: Marketing research: the role of sentiment analysis. FEP Working Paper, **489**, 1–24 (2013)

An Improved Approach to Background Removal Using Haar-Based Preprocessing for Phase Features



M. B. Veena and Meena Deshpande

Abstract The paper proposes moving object detection in Haar preprocessed images by removing background initialization constraint like a change in illumination. The fundamental objective of the proposed method is to increase the processing speed. In this work, background subtraction algorithm using phase as a feature is used to model background and distance transform is used for foreground extraction. Operation is performed with and without Haar compressed images. Experiment on real-time video streams with Haar-based phase feature implementation shows that the speedup in object detection is achieved compared to baseline method. Performance improvement of about 79% is observed in processing speed of frames with only 25% additional frame memory.

Keywords Background subtraction · Phase feature · Gabor filter · Distance transform

1 Introduction

Challenging task in computer vision application is detection of objects in changing circumstances and complex scenes like illumination variation [1]. Speed of the object varies in the temporal area. Background subtraction is a widely used method for object detection. Different methods in background subtraction are categorized into running Gaussian average, temporal median filter, mixture of Gaussians, kernel density estimation [2], co-occurrence of image variations, sequential KD approximation, eigen-backgrounds. Methods in background subtraction judge background model from the sequence of frames. Background image is the representation of the

M. B. Veena · M. Deshpande (✉)
Department of Electronics and Communication Engineering, BMS College of Engineering,
560019 Bangalore, Karnataka, India
e-mail: meenapandit26@gmail.com

M. B. Veena
e-mail: veenamb.ece@bmsce.ac.in

scene without moving objects and is to be frequently upgraded to adapt to changing luminance conditions.

There are many features used for object detection like statistical-based features which include histogram, edge [3] and corner features [4], learning-based features using intensity, SVM, texture [5]. The paper proposes phase as a feature as it is invariant to illumination changes. In this paper, background subtraction technique using phase as a feature is proposed. The proposed method is implemented on Haar compressed images (frames) to increase the speed of detection in changing illumination conditions. Gabor filter which is a linear filter is used as it can detect frequencies in specific directions. Image patch is convolved with the Gabor filter. Current value of the pixel is modeled using Gaussian distribution. Foreground detection is done by modeling phase-based background and its removal, and refinement of foreground is proposed by using distance transform. Organization of the paper is as follows: Related work is provided in Sect. 2. Section 3 describes the methodology for background subtraction using phase as a feature and distance transform. Section 4 describes current technique. Experimental results and observations are given in Sects. 4.3 and 4.4 followed by a conclusion.

2 Related Work

One of the most common methods of object detection is background subtraction, which detects moving object by subtraction of the present image pixel by pixel from the background image. Background subtraction based on per-pixel is proposed, as per-pixel method overcomes the challenges of waving trees, bootstrapping, time delay and sleeping background [6] Gaussian mixture model (GMM) [3] is a widely used background subtraction technique in the foreground detection developed in 1998. GMM is popularly employed in indoor and outdoor. Generally, three to five Gaussians are used [7]. On the basis of variance of each of the Gaussians, the foreground and background pixels are classified [8]. This technique is highly robust and used widely in background subtraction in many computer tasks. Single Gaussian mixture was developed by Wren et al. in 1997, single general Gaussian by Kim et al. in 2007, mixture of Gaussians by Stauffer and Grimson and mixture of general Gaussian (MOGG) by Allili et al. in 2007.

An adaptive GMM is used in real-time indoor and outdoor tracking in order to evaluate and determine the background of an image. Adaptive algorithm employing probability of Gaussian mixture was proposed by Zoran Zivkovic [9] to reduce the computation time, but restricting its performance in dynamic background.

The accuracy of background segmentation was enhanced by importing online learning of adaptive Gaussian mixtures on nonstationary distributions. Zoran Zivkovic [9] et al. proposed upgraded Gaussian mixture model methods and non-parametric methods [10].

Richard Palmer et al. [11] have proposed many low-level feature extraction techniques for intensity images like 2D histogram, splash features, HOG features inten-

sity and range image-based features for object detection in mobile mapping data. Texture-based features can make use of Gobor filter, which is a linear filter used to analyze the frequency content in the image in specific direction. Xue et al. [1] in the paper “background subtraction based on phase feature and distance transform” have used phase as a feature for object detection.

Sunanda R. Hanchinamani et al. suggested a method to boost the speed of operation of background detection by implementing the proposed architecture on Spartan-6 (XC6SLX45-2csg324) FPGA kit [12].

3 Background Subtraction Based on Phase Feature and Distance Transform

3.1 Basic Methodology

The work carried out by Xue, Sun and Song [1] presents a novel method for background subtraction which can work under complex environments. The proposed method works without background initialization constraints. Two stages of the proposed method are foreground detection by modeling the phase-based background and refining the foreground by the use of distance transform.

This work proposes the initial stage (of object detection) with a phase feature approach suitable for modeling background, which is done by taking every pixel being created as a group of adaptive phase features. This model captures the basic structure of foreground. Foreground detection may result in sparse pixels.

To get clear result, distance transform is used to accumulate the neighboring foreground pixels, in the second stage. In the current work, we propose some basic enhancements to this method to improve the execution time performance which forms a critical parameter.

3.2 Phase Features

Convolution of input image and local Gabor filter is performed in the proposed method. Convolution results into feature for every pixel. Each feature has many amplitudes and phase values. After convolution, every pixel has features in groups containing many amplitudes and corresponding phase values. Effective phase information per pixel in feature group is selected as per the criteria of higher amplitude. Phase information of a pixel is more accurate for a feature having higher amplitude. Sum of the selected phase features gives a new phase value.

Gaussian kernel function modulated by a sinusoidal plane wave is a 2D Gabor filter.

$$\psi_{\varphi,v}(Z) = \frac{\|k_{\varphi,v}\|^2}{\sigma^2} e^{(-\|k_{\varphi,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{\varphi,v}Z} - e^{-\sigma^2/2}]$$

where

$$\begin{aligned} \vec{k}_{\varphi,v} &= \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \phi_\varphi \\ k_v \sin \phi_\varphi \end{pmatrix}, \\ k_v &= f_{\max}/2^{v/2}, \quad \phi_\varphi = \varphi(\pi/\varphi_{\max}) \end{aligned} \quad (1)$$

where frequency $v = 0, \dots, v_{\max} - 1$ and orientation $\varphi = 0, \dots, \varphi_{\max} - 1$.

v_{\max} and φ_{\max} , respectively, depict the number of frequencies and orientations.

From Eq. (1), the first term determines oscillating part of the kernel and the second term is compensated for DC value. σ is the Gaussian window width to wavelength.

$$G_{\varphi,v}(Z) = I(Z) * \Psi_{\varphi,v}(Z) \quad (2)$$

where

* is convolution operator,

$Z(x, y)$ is the position of image,

$G_{\varphi,v}(z)$ is the outcome of convolution equivalent to the Gabor kernel at frequency v and orientation φ .

$G_{\varphi,v}(z)$, a complex value function, is expressed as:

$$G_{\varphi,v}(Z) = A_{\varphi,v}(Z) \cdot \exp(i\theta_{\varphi,v}(Z)) \quad (3)$$

where

$A_{\varphi,v}(z)$ is amplitude part and $\theta_{\varphi,v}(z)$ is a phase part $\theta_{\varphi,v}(z) \in [0, 2\pi]$.

Phase information of each pixel is extracted by dividing the image into nonoverlapping partitions, and patch-based convolution is performed. The work cites Gabor filters constructed with four frequencies and six orientations, which means $v_{\max} = 4$ and $\varphi_{\max} = 6$.

The frequency of Gabor wavelet is calculated as per formula with:

$$k_v = f_{\max}/2^{v/2}$$

with $v = 0, \dots, v_{\max} - 1$.

When $v = 3$, wavelength of the central lobe is eight pixels, capturing image local structures. Orientation is set to 6. Based on the Gabor wavelength, the patch size is selected for convolution. Based on the observed results, the choice of patch size was made between 3 and 6. Greater patch size results in rough information, and small patch size does not grab image local structures.

Based on experience, the patch size was selected equal to 4×4 . This gives different values of phase for different frequencies and orientation. This patch size gives a selective phase feature. Next task is to make a choice of the number of effective

Gabor phases for every pixel; this was done by analyzing the 4×6 convolution results.

Higher amplitude indicates larger energy, capturing correct local structure data by the filter at the selected frequency and orientation. Phase feature is built up by choosing some nominative Gabor phases and summing them together. Based on this norm, the proposal for a phase feature is made, which is built by selecting some representative Gabor phases and summing them together.

The phase feature is defined as:

$$\Phi(z) = \sum_{l=1}^L \theta_l(z) \quad (4)$$

where image position is denoted by z and feature value by $\Phi(z)$ at position z . $\theta_l(z)$ is the Gabor phase value fulfilling the condition that its corresponding amplitude is the l th highest, and L denotes the number of selected Gabor phases.

With initial experimentation using 24 phases being computed and the largest of them being selected, it was observed that beyond a certain number of the largest phase values the contribution from further ones was not considerable. Also, the further phase value range is found to be too narrow and not very stable for use for background modeling process. Thus, Gabor phases were chosen 5, $L = 5$, based on the stability and distinguishability of the new feature.

The important properties for background modeling provided by the phase feature are:

1. Insensitive to illumination changes.
2. Feature is relatively stable.
3. Wide value range.
4. Feature is discriminative.

3.3 ***Background Modeling***

Based on the phase feature, the background modeling approach for every pixel is similar and is as follows:

Over a given time interval, the phase feature of a particular pixel is called as pixel process.

$P(\Phi)$ is the probability of observing the present phase feature which is calculated as:

$$P(\Phi) = \sum_{i=1}^K \omega_i \mathcal{N}(\Phi, \mu_i, \sigma_i) \quad (5)$$

where K = no. of Gaussians,

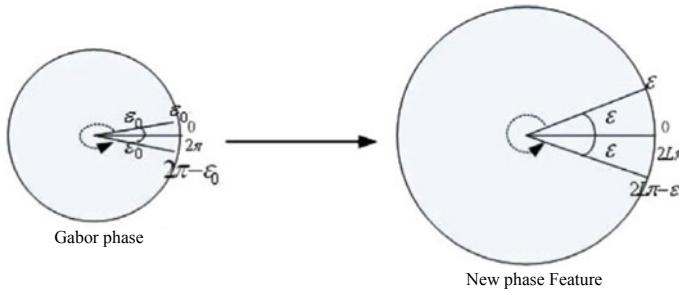


Fig. 1 Feature value warping

N = Gaussian distribution,

μ_i = mean value,

σ_i = variance of i th Gaussian,

ω_i = weight of i th Gaussian in the mixture and

$$\sum_{i=1}^K \omega_i = 1, \Phi \text{ denotes feature value.}$$

Feature U is computed and compared with the present set of K models to check for the matching models for every new pixel. The matching condition is defined as the feature value within the standard deviation of the distribution. Due to the wrapping property of the feature value at 0 and 2π , a singular region exists near the lowest and highest values. The singular region is near 0 and 2π for a given Gabor phase; for the new phase feature, the region is near 0 and $2L\pi$. Figure 1 presents the issue.

When the new phase feature and the mean value of the background model both fall into this singular area, the distance between them is computed as $2L\pi$ —Dist, where Dist is the absolute value between the two values. If such models exist, the first matching one is selected for updating its parameters and the model weights are updated as:

$$\omega_i = (1 - \alpha_1)\omega_i + \alpha_1 M_k$$

where $\alpha_1 \in (0, 1)$ is the first learning rate and for the first matching Gaussian $M_k = 1$ and is zero for others.

When the value μ_i is above the scope $[0, 2L\pi]$, the rollover to the proper side is considered, by either subtracting $2L\pi$ (if it is on higher side) or else adding $2L\pi$ if it is going below 0. If matching condition is not satisfied by any model, then the Gaussian with lowest ω is replaced by a new Gaussian with $u = \Phi$, $\sigma = \sigma_{\text{init}}$ and an initial weight.

3.4 Blob Aggregation Using Distance Transform

Descending order is used to sort the Gaussian in foreground detection stage, according to values of x/r ratio.

First B distributions are chosen as the background.

$$B = \arg \min \left(\sum_{i=1}^b \omega_i > T \right) \quad (6)$$

Minimum probability of the background to be part of the scene is represented by the threshold value T . Feature value of the pixel is compared with the B distribution, and the pixel is classified as background if the feature value matches with B distribution; else, it is classified as foreground.

Distance map is the depiction, where each pixel has a value indicating its distance from the nearest white pixel. Distance transform is used to convert a binary image into a distance map. Distance value is zero for one pixel value.

A $M \times M$ binary image can be represented by

$A = \{(i, j): a(i, j) = 0 \text{ or } 1\}$, for $i, j = 1, \dots, M$, and

$W = \{(x, y): w(x, y) = 1\}$ are the coordinates of white pixels in the image.

The Euclidian distance transform for a pixel $a(i, j)$ is given as:

$$d(i, j) = \min_{(x, y) \in W} \sqrt{(i - x)^2 + (j - y)^2} \quad (7)$$

Pixels with lower values correspond to foreground and are compact in the distance image, while white pixels (Fig. 2) Query focuses on the foreground region. Distance image is further segmented with a threshold value. The good value for threshold is found to be between 1.2 and 1.5. Proper choice of threshold value preserves the foreground structure and suppresses the noise. Noise is further removed by morphological operations. Figure 3 presents the impact of blob aggregation.

The main steps in blob aggregation are:

- (a) A dilation operation radius is performed on the initial detection results by using disk-shaped structuring element with adjustable radius.
- (b) Distance map is segmented with threshold value 1.2 by applying distance transform to get binary image.
- (c) Segmented results containing connected areas having pixels fewer than 200 are removed.
- (d) Holes of the binary image are filled to complete foreground mask. Dilation operation using disk-shaped structuring element is performed again before filling gaps to get closed boundaries.
- (e) Final result is obtained after applying erosion operation to thin the detected mask. Disk-shaped structuring element with double adaptive radius is used for erosion operation.

Figure 2 presents the complete flow of the baseline method.

```

construct the Gabor filter bank, set the image patch size.
initialize the phase based background model

For each new frame
    do the patch based Gabor filter convolution
    For(each pixel in the frame)
        compare its phase feature  $\phi$  with existing models
        If (matched model is found)
            | update the matched model parameters
        Else
            | the model with lowest weight is replaced
        End
        sort the background models according to  $\omega/\sigma$ 
        select the first B distributions as background model
        If (feature  $\phi$  fits at least one background model)
            | Cur_Pixel = background
        Else
            | Cur_Pixel = foreground
        End
    End
    blobs aggregation using distance transform and
    morphological operations
End

```

Fig. 2 Baseline method flow

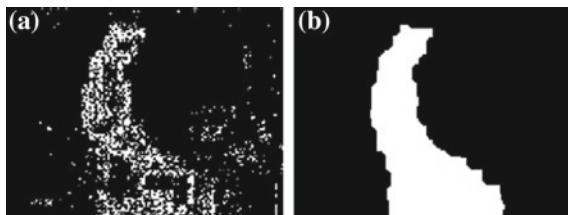


Fig. 3 **a** Impact before blob aggregation. **b** Impact after blob aggregation

4 Proposal of Current Work

4.1 *Background*

The baseline method operates on video frames starting from frame 1 onward, and each frame is processed sequentially. As the computation of Gabor convolutions as well as the selection of the phase feature values is computationally resource consuming in terms of memory as well as processing speed, this work makes an attempt to improve the processing speed of the baseline method.

4.2 *Technique*

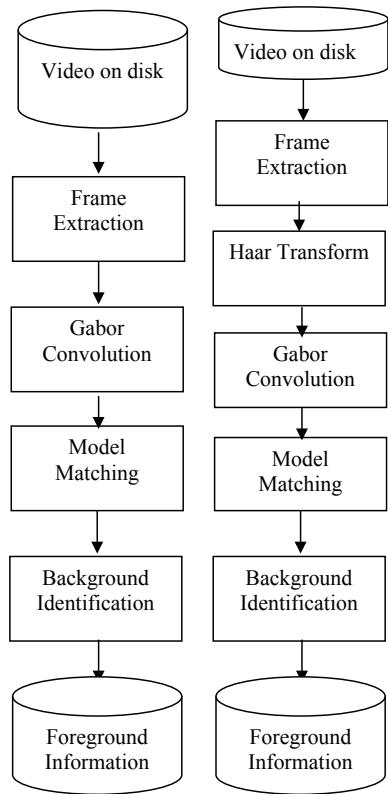
The processing time for the baseline method is seen to be split into the following:

1. The time component consumed by one-time initialization
2. The time used for per frame processing which depends on the following parameters:
 - a. Total number of video frames
 - b. Width of the video frame
 - c. Height of the video frame
 - d. Pixel color data structure.

The speedup can be achieved by reducing any of these parameters. Say for a simple speedup, one can try to skip alternate video frame, only process every other frame using the baseline method and get a speed advantage near to the double of original speed. This is an obvious approach, but it surely will tend to lot of noisy detections of the objects as the phase features will be having a lot of variations due to the discontinuity in the input image frames being increased due to dropping alternate frames. The other speedups can be achieved by using a pixel color data structure and work on it, but it will be at other end of the spectrum in terms of improving performance (Fig. 4).

A better approach would be to process every frame but not try to look at individual pixel color data structure and try to minimize the processing using some methodology. The current work exactly targets this approach and makes use of a transform for reduction of data to process and hence have a reduction in the processing activity. In order to carry out this, we use a simple approach of involving a simple wavelet—Haar wavelet for the purpose. It works like a compression operation, thereby reducing the pixel count to process at each frame level to one-fourth of the original, and provides a boost in the processing speed.

Fig. 4 Baseline **a** and current proposal **b** flow



4.3 Experimentation

In this work, we exploit the content available from approximation area after performing Haar transform-based process on each frame for improving the processing time per frame. As the per frame time improves, the overall time for video processing timing improves. As the background removal process is done on the approximation subband image while it is having half width and half height as compared to the input video frame the bounding box which is put up for the object is of scaled size (i.e. having half width and half height) and also its top left (the anchor point) is located half the distance from the first pixel along both axes (Table 1).

In order to locate the object in the original video, it is needed to transform these bounding box coordinates to the proper position relative to the input frame from the video. For this purpose, the bounding box coordinate is obtained from Haar-based process which needs to be scaled by 2 and is used for marking the object. So, the object position in the input frame is declared at $2 * [Left, Top, width, Height]$.

In order to establish the fact that the use of Haar-based approximation sub-band helps in the performance improvement, we have processed multiple videos using

Table 1 Comparative performance against baseline

Frame no.	1	2	3	4	5	Average	Total
Baseline	97.98	52.92	54.30	52.25	53.38	62.17	310.83
Haar-based	13.86	12.72	12.52	12.43	12.42	12.79	63.94
% Time	14.14	24.04	23.05	23.79	23.27	20.57	20.57
Saved time (%)	85.86	75.96	76.95	76.21	76.73	79.43	79.43

both the baseline and the modified Haar transform-based approach. MATLAB was used as a tool as the implementation was done in software. A laptop system with processor 1.70 GHz and RAM running at 8 GHz was used for the execution. The time for processing per frame was recorded in both cases; also, the memory used up at the end of the script was listed using “whos” command available in MATLAB.

4.4 Observations

A sample reading for five frames of a video is provided herein from the log which depicts the processing time reduction capability of the proposed approach.

Timing for the video frames using baseline method

Frame-Level Timing

97.9817 52.9182 54.2992 52.2477 53.3783

Total Processing Time = 310.8251 s

Timing for the video frames using Haar-based method

Frame-Level Timing

13.8561 12.7215 12.5160 12.4286 12.4221

Total Processing Time = 63.9443 s

The observation here is as follows:

1. The time needed for the first frame processing is high as compared to all subsequent frames in baseline as well as proposed method.
2. In every frame, the time needed to process for Haar-based method is less than 25% of that needed by the baseline method.
3. The average time for processing a single frame for the baseline method is around 62.17 s and that of Haar is 12.79 s, a reduction of about 79.43%.
4. The high time for the first frame can be possibly due to the initial memory allocation needed for the process (Fig. 5).
5. In every frame, the time needed to process for Haar-based method is less than 25% of that needed by the baseline method.
6. The average time for processing a single frame for the baseline method is around 62.17 s and that of Haar is 12.79 s, a reduction of about 79.43%.

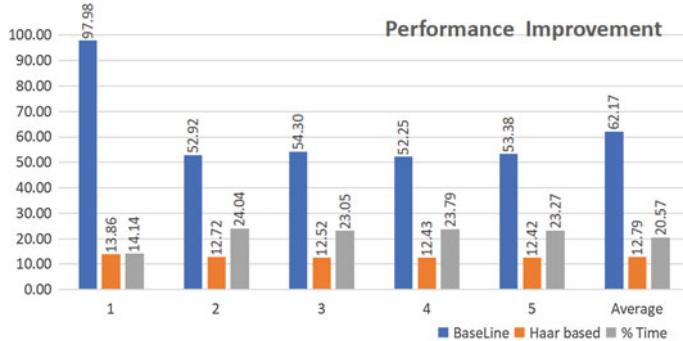


Fig. 5 Performance improvement using Haar-based technique

7. The high time for the first frame can be possibly due to the initial memory allocation needed for the process.

For the purpose of computations of the Haar transform, each of the frames from video which we call as current frame $N \times N$ is processed using Haar transform and four frames of its half height ($N/2$) and width ($N/2$) are generated. This leads to the following additional memory requirements for a video with frames of size 640×320 (RGB).

It is seen that a total of five frames are needed of $(N/2) \times (N/2)$, one holds the Haar frame to process, and the rest hold the four sub-bands resulting from the Haar. Thus including variables to record the Haar process, it is observed that a total of 2,304,000 bytes are needed which amounts to $2,304,000/1,843,200 = 1.25$ frames. The Haar transform-based approach needs only 25% more memory.

5 Conclusion

In this work, an object detection using background subtraction method based on Haar transform-based approach to speedup phase feature-based technique has been proposed. It is observed that the Haar-based approach needs a very low additional memory requirement of 25% of a single frame memory and provides a performance gain which saves about 79% time on average basis while processing the video frames for background removal as compared to the baseline method (Fig. 6 and Table 2).

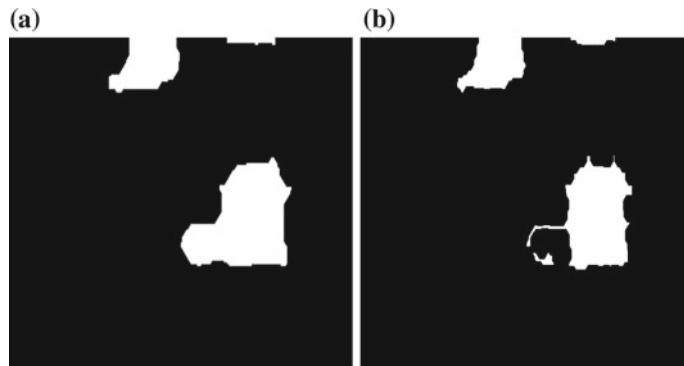


Fig. 6 **a** Object detection using baseline method. **b** Object detection using baseline method on Haar transformed image

Table 2 Additional memory requirements for Haar process

Object in process	Memory needed	Memory in bytes
CurrentFrameGray (for both methods)	360×640	1,843,200 (double)
cAintCurrentFrameGray	180×320	460,800 (double)
cHintCurrentFrameGray	180×320	460,800 (double)
cVintCurrentFrameGray	180×320	460,800 (double)
cDintCurrentFrameGray	180×320	460,800 (double)
CurrentFrameGrayHaar	180×320	460,800 (double)
	Total	2,304,000 bytes

References

1. Xue, G., Sun, J., Song, L.: Background subtraction based on phase feature and distance transform Elsevier B.V (2012)
2. Elgammal, A., Duraiswami, R., Harwood, D., Davis L.S.: Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. In: Proceedings of the IEEE **90**(7), 1151–1163
3. Dikmen, M., Huang, T.S.: Robust estimation of foreground in surveillance videos by sparse error estimation. In: Proceedings of the 19th international conference on pattern recognition, pp. 1–4. (2008)
4. Saif, A.F.M.S., Prabuwono, A.S., Mahayuddin, Z.R.: Moment feature based fast feature extraction algorithm for moving object detection using aerial images. PLoS ONE **10**(6), e0126212 (2015) <https://doi.org/10.1371/journal.pone.0126212>
5. Heikkilä, M., Pietikäinen, M.: A texture-based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)
6. Heikkilä, M., Pietikäinen, M.: A texture-based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)
7. Wren C.R., Azarhayejani, A., Darrell, T., Pentland, A.P.: Pfinder: real-time tracking of the human body. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 780–785 (1997)

8. Cristani, M., Farenzena, M., Bloisi, D., Murino, V.: Background subtraction for automated multisensor surveillance: A comprehensive review. *EURASIP J. Adv. Signal Process.* 1–24 (2010)
9. Zivkovic, Z., van der Heijden, F.: Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recogn. Lett.* **27**(7), 773–780 (2006)
10. Zhang, Jianshu, Cao, Jie, Mao, Bo: Moving object detection based on non-parametric methods and frame difference for traceability video analysis. *Procedia Comput. Sci.* **91**, 995–1000 (2016)
11. Palmer, R., Borck, M., West, G., Tan T.: Intensity and range image based features for object detection in mobile mapping data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXIX–B3 2012 XXII ISPRS Congress, Melbourne, Australia 25 Aug–01 Sept (2012)
12. Hanchinamani, S.R., Sarkar, S., Bhairannawar, S.S.: Design and implementation of high speed background subtraction algorithm for moving object detection. In: 6th international conference on advances in computing & communications, ICACC, Cochin, India, 6–8 Sept 2016

Artificial Neural Network-Based Crop Yield Prediction Using NDVI, SPI, VCI Feature Vectors



Preeti Tiwari and Piyush Shukla

Abstract Agriculture is one of the major revenue producing sectors of India and a source of survival. The number of biological, financial, and environmental factors affects the crop yield. So, unidentifiable changes in these factors lead to failure in agriculture area. This paper focuses on prediction of crop yield where different geospatial features were utilized, such as normalized difference vegetation index, standard precipitation index, vegetation condition index. In order to learn from previous weather condition, a standard error back propagation neural network was used. Here, the training was done in such a way that all set of features were utilized in pair with their yield value as output. For increasing the reliability of the work, whole experiment was done on real geo-spatial dataset from Madhya Pradesh region of India. Result shows that proposed model has overcome various evaluation parameters on different scale as compared to previous approaches adopted by researchers.

Keywords Crop yield prediction · Data mining · Machine learning · Vegetation index

1 Introduction

In past decades, IT has become more and more part of our everyday life. So, this Information and Technology upgrades productivity for any piece of industry and administrations get enhanced; this is particularly valid for agriculture as well. Nowadays, a farmer not only reaps or trims its crop but helps in developing measures of information. This information is exact and small in scale.

However, gathering a lot of information frequently is both a gift and a revile. There is a considerable measure of information accessible containing data about

P. Tiwari · P. Shukla
UIT, RGPV, Bhopal, India
e-mail: dubeypreeti@yahoo.com

P. Shukla
e-mail: piyush@rgtu.net

certain advantage. Here soil and yield properties, which ought to be utilized for the farmer's advantage. This is a typical issue for which the term information mining has been involved. Information mining systems go for finding those examples or data in the information that are both important and fascinating to the farmer [1–4]. A typical particular issue that happens is crop yield recommendation. Previously, this yield forecast has really depended on agriculturist's long-term understanding for particular yield, crops, and climatic conditions. Be that as it may, this learning may likewise be accessible, yet covered up in the small scale. Exact information which would now be able to days gathered in seasons utilizing an enormous measure of seasons [5, 6].

Updated and balancing information gives out the rural generation at a quicker pace for rural improvement. Productions of any harvest lead either by consideration of zone or change in profitability or both. In India, the likelihood of broadening the territory under any product, nearly, does not exist with the exception of by reestablishing to expand trimming power or harvest substitution [7–9].

2 Related Work

In [1], Pritam Bose et al. outline this idea with the presentation of the new type of neural network in crop yield prediction field of research name as spiking neural network where NDVI feature vector was used which was sequence of time series images. Here, spatial dataset images of NDVI were utilized for the training of spiking neural network. Here, the selection of some blocks was done in the training, so this increases the input data quality which increases the accuracy of prediction. The SNN approach was connected to estimate the wheat yield in Shandong region which was one of the principle wheat-developing regions. This technique could forecast the yield around a month and a half before harvest with a high precision values.

In [10] Michele Meroni et al. displayed the present investigation supplements this quality evaluation by breaking down the impact of utilizing PV NDVI rather than VGT NDVI for operational product checking and yield anticipating work. The research mentions the utilization of combined objective facts from the instruments gathered during the time frame when both satellite frameworks were dynamic.

In [11], Gandhi et al. displayed the review on use of machine learning framework for Indian rice altering ranges.

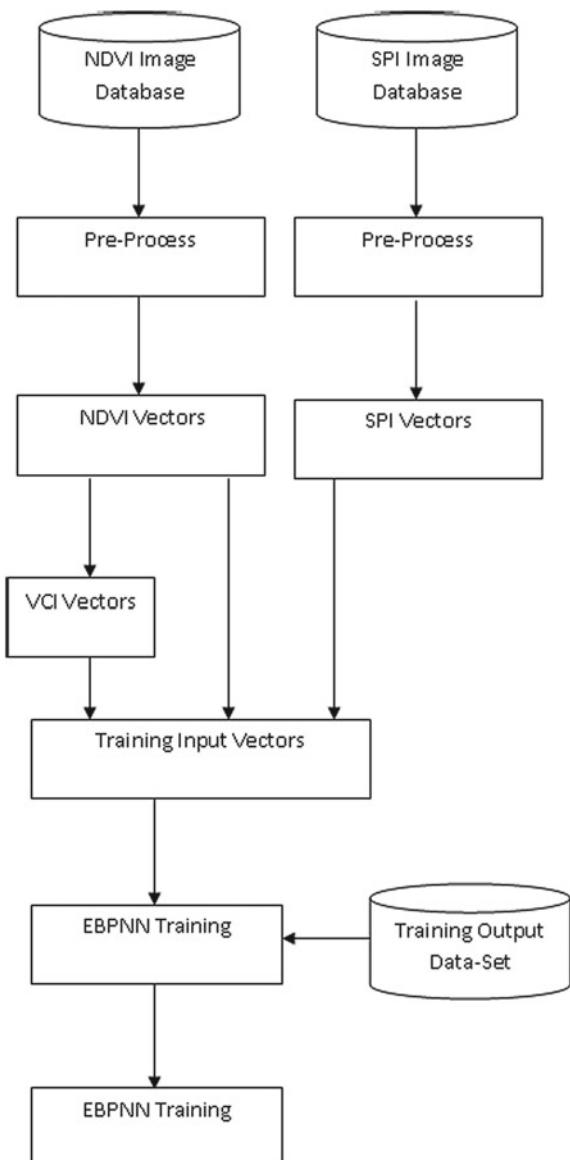
In [12], Geetha inspected about the snippet of data mining in setting of developing field and besides introduces around a couple of data mining frameworks and their related work by a couple of makers in setting to cultivation territory.

Agriculture is the most imperative application, and the use of information advancement association over agriculture may change those conditions from asserting decision making. Furthermore, agriculturists could yield secured nearby an uncommon approach. Data mining accepts a critical part to decision making investigating a couple of issues related to cultivating field.

3 Proposed Methodology

In this work, the use of different vegetation indices was done as a feature vectors. Explanation of each was done in this section with their requirement and formulas, and the proposed work is shown with the help of block diagram shown in Fig. 1.

Fig. 1 Block diagram of proposed work EBPNN for crop yield prediction



3.1 Normalized Difference Vegetation Index (NDVI)

For crop yield prediction, NDVI is ordinarily utilized as an input feature vector. This feature measures vegetation cover in the land. NDVI was first recommended as a record of vegetation well-being and thickness. It is as curtained in Eq. 1.

$$N = \left(\frac{b_{\text{NIR}} - b_{\text{RED}}}{b_{\text{NIR}} + b_{\text{RED}}} \right) \quad (1)$$

where variable N represents as the normalized difference vegetation index (NDVI), while b_{RED} are the reflectance of red bands and b_{NIR} stands for near infrared.

3.2 Standardized Precipitation Index (SPI)

In 1993, this index was introduced by McKee and his team member. SPI helps in quantifying the precipitation shortage occurs in different time scale for reflecting the effect of available water resources. So, temporal pattern was identifying by estimating the standard precipitation value for different month period. In order to estimate SPI value, difference of the precipitation from the average of selected time scale was taken and finally divides it by the standard deviation.

3.3 Vegetation Condition Index (VCI)

It is suggested in [3], which represents correlation between the current NDVI value with the minimum NDVI value from a huge set of period of time. It was calculated as shown in Eq. 2, where V_j stands or the j th position VCI value in the vector. Similarly, N_j stands for the j th position NDVI value in the vector. However, N_{\min} and N_{\max} are minimum and maximum values of NDVI vector.

$$V_j = \left(\frac{N_j - N_{\min}}{N_{\max} - N_{\min}} \right) \times 100 \quad (2)$$

3.4 Preprocessing

Input image data need to be preprocessed first for transforming data as per proposed model. It involves activities like cleaning of data and convert in required environment format. This can be understood as the let image consist of four pixel having dimension of 2×2 then for this image a matrix is of same dimension 2×2 and its four cell contain value as per the pixel color and representing format. In this step, image is

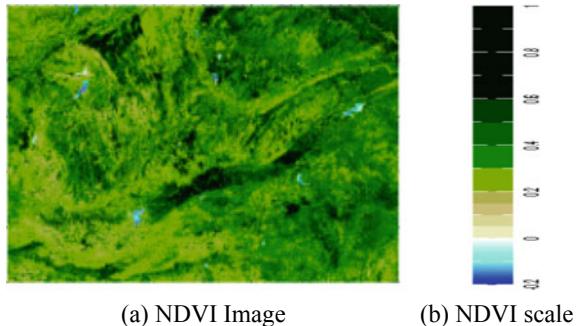
resized in fix dimension. As different image have different dimension, the conversion of each is done in this step.

3.5 Index Vector

In order to get index value from the image, each image has its own color scale where pixel value corresponding to a specific color is replaced by its index value. This can be understood as Fig. 2a has NDVI index than all the dark green portion in the image is consider as the extremely wet region so if this dark green is represent by pixel value 65 than all the pixel value in the image of NDVI where 65 value is present is consider as extremely wet region and NDVI value for the cell is +1.0. In similar fashion, other regions are also detected in the image. However, NDVI image index value is obtained from its separate index as image color region of this image is different.

In this work, three indexes are use as the training parameters first was NDVI, second was SPI, and third was VCI. The use of three feature vector for training increases the accuracy of the resultant neural network. This helps in covering various other geographical parameters of the location. So, a set of values for any geographical location at particular period of time, this feature vector have three values set. This can be understood as if geographical location (x_i, y_i) than its corresponding feature vector for time instance t_i is $[N_i, V_i, S_i]$. While for same location but at time t_j , it might be $[N_j, V_j, S_j]$. So, the data input feature vector for each year is three dimension vector, where first and second dimension provides geographical position while third provide index vector for the same.

Fig. 2 Input data for getting NDVI vector



3.6 Training Vector

Here, any consecutive term set is considered as the pattern in the document. As it is known that collection of patterns is performed in the separate set of features, input and output parameters are combined to generate a training data for the error back propagation neural network.

3.7 Training of Error Back Propagation Neural Network (EBPNN)

- Let us assume a three-layer neural network.
- Now, consider i as the input layer of the network, while j is considered as the hidden layer of the network.
- Finally, k is considered as the output layer of the network.
- If W_{ij} represents weight of the between nodes of different consecutive layers.

So, the output of the neural network is depending on the below Eq. 3:

$$Y_j = \frac{1}{1 + e^{-x_j}} \quad (3)$$

where $X_j = \sum x_i \cdot W_{ij} - \theta_j$, $1 \leq i \leq n$; n is the number of inputs to node j , and θ_j is threshold for node j . Let input vector of three values which include values from NDI, VCI, SPI images passes from the first layer of neural network. These values get multiply by above weight values. Now, this acts as input H_{input} to next layer of hidden neurons. In this, some biasing is also possible, which was neglected in this example. So, the weight values of the neuron for next level were assumed. Where, each value obtained from the previous weight matrix multiplication is passed through the sigmoidal function 4. Therefore, small variation in the output value was done by this function. Let the expected output vector be $y = [0 \ 1 \ 0]$ and obtained values from output layer was $O = [0.6536, 0.7197, 0.6761]$, and the difference between the expected with obtained is consider as the error. This error needs to be corrected by adjusting the weight values of each layer. So here, the forward movement of the neural network is over and error back propagation starts by Eq. 5.

$$\frac{\partial E_i}{\partial O_i} = \frac{\partial (-1 * ((y_i * \log(O_i)) + (1 - y_i) * \log(1 - O_i))}{\partial O_i} \quad (4)$$

$$\frac{\partial E_i}{\partial O_i} = (-1 * (y_i * \log(O_i)) + (1 - y_i) * \log(1 - O_i)) \quad (5)$$

In similar fashion, other values can be calculated to find other set of derivatives for sigmoid of Eq. 6. Here as per output derivative, value may vary.

$$\text{Sigmoid} = (1/(1 + e^{-x})) \quad (6)$$

$$\frac{\partial O_i}{\partial H_i} = \frac{\partial(\frac{1}{1+e^{-x}})}{\partial x} \quad (7)$$

$$= ((1/(1 + e^{-x}) \times (1 - (1/(1 + e^{-x})))) \quad (8)$$

For each input to neuron, the derivative with respect to each weight is calculated. Now, let us look at the final derivative:

$$\sum_{i=1:n} \frac{\partial H_i}{\partial W_{i(j,k)}} = \frac{\partial(h_{i(\text{output})} * W_{i(j,k)})}{\partial W_{i(j,k)}} \quad (9)$$

Now, by using chain rule final derivates were calculated for the same. Here, multiplication of output obtained from Eqs. 5, 6 and 7 was done in following way:

$$\frac{\partial E_i}{\partial W_i} = \frac{\partial E_i}{\partial O_i} * \frac{\partial O_i}{\partial H_i} * \frac{\partial H_i}{\partial W_i} \quad (10)$$

So, overall ∂W_i can be obtained by getting value of weight from above equation; here, all set of weights which need to be update are change by below matrix values.

- So, error corresponding to the input data was estimated by differencing desired output obtained from output layer.

$$e_k(n) = d_k(n) - y_k(n) \quad (11)$$

- The EBPNN weight updation was done by matrix of ∂W_i and Eq. 12.

$$W_{ij} = W_{ij} + \Delta W_{ij} \quad (12)$$

- So end of above iteration, steps over when error obtained from the output layer get nearer to zero or some constant such as 0.001.

3.8 Testing of EBPNN

Testing of trained neural network obtained from above steps is processed by passing testing input images of NDVI, SPI. These images are preprocessed and generate feature values of as done in training part of the neural network. Finally, feature vectors were pass in the EBPNN which gives crop yield predicted value. This crop yield predicted values are compared on various parameters to find the fitness of the proposed work.

4 Experiment and Results

In order to conduct experiment and measure evaluation results, MATLAB 2012a version software is used. This section of paper shows experimental setup and results. Experiment was done on machine having configuration of 4 GB RAM, with sixth generation Intel I3 processor.

4.1 Dataset

Experiment is done on real dataset having values of the SPI from <https://iridl.ldeo.columbia.edu/maproom/Global/Precipitation/SPI.html>. While NDVI value obtained from <http://iridl.ldeo.columbia.edu/SOURCES>. Each source image is extract on the longitude parameters 21°–26°N 60, and latitude parameters 74E–82E, where selected region is one of state Madhya Pradesh in India. However, average of different crop yields is obtained from <http://www.mospi.gov.in/statistical-year-book-india/2016/177>.

4.2 Results

Results are compared with the previous work spiking neural network (SNN) in [1] which is termed as previous work in this paper.

It has been observed by Table 1 that relative error of proposed work of crop yield prediction work is less as compared to the previous method adopted in [1]. One more important factor is observed that with the change in climatic condition, the production of crops also raises.

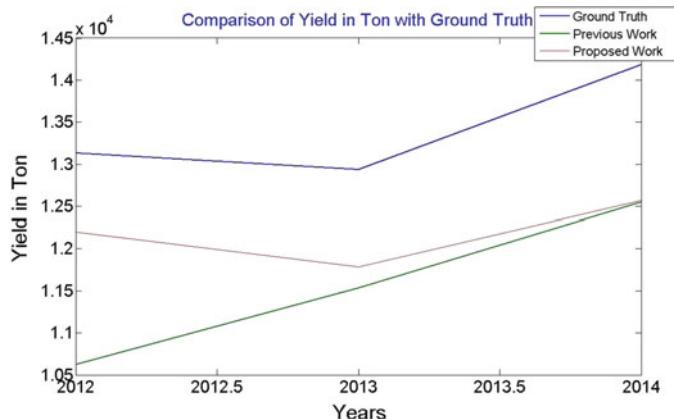
It has been observed by Table 2 and Fig. 3 that training time and relative error in error back propagation neural network are comparatively low because of less number of neuron in input layer as compared to spiking neural network used in the previous work [1]. Here, input layer in the EBPNN is dependent on number of feature used, while in case of spiking neural network, a number of neuron in input layer depend on training block size.

Table 1 Relative error value comparison of previous work with proposed work

Prediction year	Relative error		
	Previous work	Proposed work	Improved efficiency %
2012	19.1075	7.1858	62.4
2013	10.8299	9.1835	15.2
2014	11.5063	1.3895	87.92

Table 2 Training time value comparison of previous and proposed work

Iterations	Training time in seconds		
	Previous work	Proposed work	Improved efficiency %
10	29.36	15.23	48.1267
20	42.58	26.98	36.6369
40	74.46	48.23	35.2269

**Fig. 3** Yield in ton value comparison of previous work with proposed work**Table 3** RMSE value comparison of previous work with the proposed work

Prediction Year	RMSE		
	Previous work	Proposed work	Improved efficiency %
2012	2509.4	943.7	62.3934
2013	1401.1	1188.1	15.2023
2014	1615.3	1611.8	1.02148

It has been observed by Table 3 that proposed work of crop yield prediction using multiple features works well as compare to the previous method adopt in [1]. Here, RMSE value of proposed work is low as compared to spiking neural network model. In this work, training feature vector increases the value of accuracy as well.

5 Conclusion

As digital data on libraries and servers are drastically increased with every new second, researchers are attracted to work on it. So, this work focuses on geo-spatial data for crop yield prediction by utilizing various features vector of SPI, NDVI, and

VCI. Classification based on working of neural networks has also been performed by many researchers. In some of the works, document learning is based on single information, but through this proposed work it will overcome this dependency, as well as the learning can be performed on all the available data. Here, the result shows that proposed work has improve the prediction accuracy by reducing the RMSE, relative error value. Also through proper training and rich input vector resultant, the neural network is less time-consuming model. It was obtained that proposed work has reduced the relative error by 33% as compared to SNN, while RMSE also got reduced by 31.8%. In this work, average testing time for prediction of crop yield was decreased by 38.2% and overall accuracy was improved.

References

1. Bose, P., Kasabov, N.K., Bruzzone, L., Hartono, R.N.: Spiking neural networks for crop yield estimation based on spatiotemporal analysis of image time series. *IEEE Transactions on Geoscience and Remote Sensing* (2016)
2. Ramesh, D., Vardhan, B.: Analysis of crop yield prediction using data mining techniques. *Int. J. Res. Eng. Technol.* **4**(1), 47–473 (2015)
3. Mr. Mankar, AB., Mr. Burange, MS.: Data mining—an evolutionary view of agriculture. *Int. J. Appl. Innov. Eng. Manage. (IJAIEM)*, **3**(3), 102–105 (2014)
4. Raorane, A.A., Kulkarni, R.V.: Review-role of data mining in agriculture. *Int. J. Comput. Sci. Info. Technol.* **4**(2), 270–272 (2013)
5. E. Manjula, S. Djodiltachoumny “A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017 ISSN: 2349–6363 298
6. Grajales, D.F.P., Mosquera, G.J.A, Mejia, F., Piedrahita, L.C., Basurto, C., “Crop-Planning, Making Smarter Agriculture With Climate Data”, Fourth International Conference on Agro-GeoInformatics, pp. 240–244, 2015
7. Medar, R., Rajpurohit, V.: A survey on data mining techniques for crop yield prediction. *Int. J. Adv. Res. Comput. Sci. Manage. Stud.* **2**(9), 59–64 (2014)
8. Kalpana, R., Shanti, N., Arumugam, S.: A survey on data mining techniques in Agriculture. *Int. J. Adv. Comput. Sci. Technol.* **3**(8), 426–431 (2014)
9. Fatima, G.N., Geeta, R.: Agriculture crop pattern using data mining techniques. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **4**(5), 781–786 (2014)
10. Meroni, M., Fasbender, D., Balaghi, R., Dali, M., Haffani, M., Haythem, I., Hooker, J., Lahlou, M., Lopez-Lozano, R., Mahyou, H., Moussa, M.B., Sghaier, N., Wafa, T., Leo O.: Evaluating NDVI data continuity between SPOT-VEGETATION and PROBA-V missions for operational yield forecasting in North African countries. *IEEE Trans. Geosci. Remote Sensing* **54**(2), 795 (2016)
11. Gandhi, N., Armstrong, L.J., Petkar, O., Tripathy, A.: Rice crop yield prediction in india using support vector machines. In: IEEE The 13th international joint conference on computer science and software engineering (JCSSE), Thailand (2016)
12. Geeta, M.C.S.: A survey on data mining techniques in agriculture. *Int. J. Innov. Res. Comput. Commun. Eng.* **3**(2), 887–892 (2015)

A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT



Ashwani Kumar

Abstract The authentic document should be available for the legitimate user so that none other than him can use the document; there are some instances to protect any document from modification, or distribution of the properties in non-authentic way. In this, author has reviewed and implemented the discrete wavelet transform and singular value decomposition with LSB-based techniques to protect the digital rights. The schemes is to develop authentic intellectual properties and prevent them from exhibiting to unknown user. The cover image (the image in which watermark is inserted) is to be fractioned into (HH, LH, LL, and HL) by using commonly used method of transforming frequency domain. This transform can be directly applied to each and every subordinate bands of the original image. The extraction mechanisms work in accordance with the previous work done by the embedding algorithm. The output and the frequency of image will determine strength, apparently maintaining the better quality and robustness to many attacks to compromise the security of intellectual property such as JPEG compression, Gaussian noise, and there are many more.

Keywords Intellectual properties · Image compression · Frequency distribution

1 Introduction

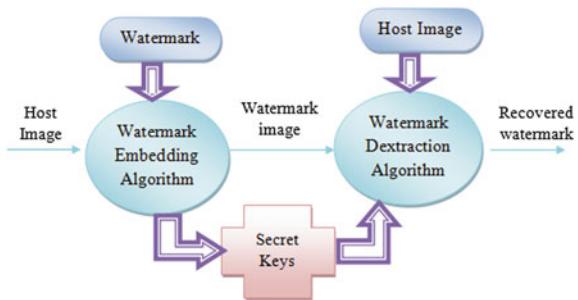
The techniques of digital image watermarking [1, 2] aim to embed a secretive signal known as watermark into an image. There are certain application areas of digital image watermarking such as protection of the intellectual properties, monitoring, detecting the tamper information in the property, right document mark, verifying the integrity, description, fingerprinting, and protection of channels at receiver and sender. The content protection [3] can be accomplished by using the following ser-

A. Kumar (✉)

Department of Computer Science Engineering, Vardhaman College of Engineering, Shamshabad, Hyderabad, Telangana, India

e-mail: ashwani.kumarcse@vardhaman.org

Fig. 1 Demonstrates process took place in watermarking



vices: document authentication, verifying any one that is claiming the right on that document, to protect data from unauthorized disclosure, i.e., confidentiality; the next is data integrity that aims to check that certain bits have not been altered, damaged, or lost in an accidental manner or an authorized manner.

The schemes [4] can be utilized for the encryption by including secret key in the sender's message so that the originality of the sender's message can be preserved. Figure 1 explains the basic idea behind the process that took place in authenticating a document by using digital image watermarking.

The insertion of a secret key that is unknown (not public) in the sender's message is known as embedding process of watermarking. The schemes [5] can determine whether the integrity and confidentiality of the intellectual property are preserved or not. However, there is a problem with these schemes, that is, they cannot simplify the coordinates where the problem has been occurred exactly.

2 Related Work

In this section, the background of digital image watermarking techniques such as LSB-based watermarking methods and discrete wavelet transform (DWT)-based methods is shown.

2.1 *LSB Technique*

Now for a digital image, schemes [6] can insert any bit of data at desired coordinate, and also, there can be a trick that can be utilized to hide data in the plain sight where the area is so denser in the cover image so that hiding of secret information can be done without much amendments to other areas of the image. Rastegar and Bruce [7] proposed very old schemes that were used when people identified the concept of hiding in the plain sight, and these methods were proposed for two kind of services or key idea to implement digital image watermarking. These methods [8] were not

Fig. 2 Three-level DWT decomposition

LL ₃	HL ₃	HL ₂	HL ₁
LH ₃	HH ₃		
LH ₂		HH ₂	
LH ₁		HH ₁	

that complex as DWT and DCT. By changing bits in order to a particular image can show results and since it was based on changing least bit in the image thus this method is called LSB. The techniques given by Kurah to implement the above method proposed were downgrading. These methods were easy to utilize and also have better significance.

2.2 DWT Technique

It is a commonly used technique for the watermarking that involves the transformation of the image from the particular pixel domain to certain frequency domain. It has been proved that the wavelet coefficient [9–14] has an upper hand than most of the widely used algorithms than the other scheme. It is a method that breaks any one-dimensional signal into two parts known as detail coefficient and approximation through decomposition process. The high and low frequencies of the signal are analyzed via high-pass filters and low-pass filters in DWT. The results of such filters are referred to as DWT coefficients, and with the help of the coefficients, the original watermark image can be reconstructed. This entire process of reconstruction is known as inverse DWT (IDWT). Figure 2 represents three levels of DWT decomposition.

3 Implementation of Digital Image Watermarking Using LSB and DWT

In this hybrid watermarking scheme, the author has used the least significant bits and wavelets for better safety of digital content during the transmission of it over non-secure channel. The scheme ensures [15] that the hybrid scheme is accepted in terms of all basic image-processing operations.

3.1 LSB Procedure

Here, the author has shown embedding procedure of hybrid watermarking scheme using the least significant bit (LSB) procedure. The steps given below are taken to embed the watermark into the cover object using the secret key encryption algorithm.

Step 1: Input 1: Original image.

Step 2: Input 2: Watermarked image.

Step 3: Replace the least significant bit of cover image by most significant bit of watermarked image.

Step 4: Output 1: Original image

Output 2: Watermarked image.

Step 5: Attack different types of noises with different variances.

Step 6: To find watermarked and original image, apply Inverse of LSB.

In the LSB watermarking, a watermarked image is inserted into the primary image. After the completion of watermarking process, the watermark data are extracted from the cover object.

3.2 DWT Procedure

In this section, the embedding procedure of hybrid watermarking scheme in discrete wavelet domain is shown. The steps given below are taken to embed the watermark in the cover object.

Step 1: In this step, three-level DWT is computed to the original object to break down into LL₃, LH₃, HL₃, and HH₃ subgroups.

Step 2: After applying three-level DWT, the singular values are chosen for the clamorous watermarked image, \mathbf{W}_{ACM} .

Step 3: In this step, author has calculated SVD for all high-frequency subgroups of the original work (LL₃, LH₃, HL₃, HH₃).

Step 4: Change the singular values \mathbf{S}^i with \mathbf{S}_{ACM} of chaotic watermarked image, where gaining factor is fixed for all subgroups and i represents remaining groups.

Step 5: Applying inverse SVD upon the original image that was transformed using some modified values.

Step 6: In the last step, calculate inverse three-level DWT using the updated coefficients to develop the watermarked work.

4 Result and Analysis

The various performance metrics are used to calculate the robustness and imperceptibility.

Mean Square Error: It is used to calculate the average of the squares of the errors. The value of that parameter closer to zero is better.

$$\text{MSE} = \sum \frac{\sum (Y - \hat{Y})}{m \times n}$$

Peak Signal to Noise Ratio: PSNR is used to analyze the quality of image. A greater value of that parameter means quality of the picture is high.

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_f}{\sqrt{\text{MSE}}} \right)$$

In this experiment, watermark is individually inserted into the LL₃, LH₃, HL₃, HH₃ sub-bands of the original image.

Figure 3 shows the actual implementation of the approach, and in Fig. 4, the image on the right-hand side is the watermarked image which is found after inserting watermark, beside that the watermark image which is to be inserted.

Figure 5 shows the extracted image that was embedded in the process, and if it is the same image that was inserted, then that means authenticity is verified. Here, the author has demonstrated the performance of hybrid watermarking scheme under

Fig. 3 LSB-DWT-based watermarking

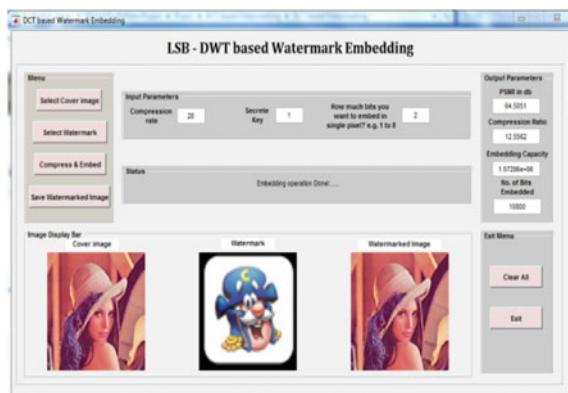
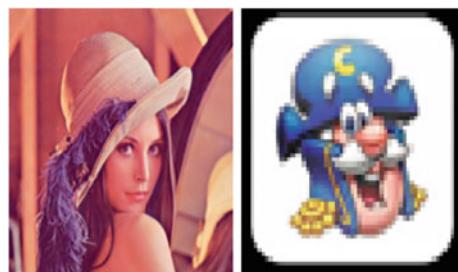


Fig. 4 Original image and watermark image



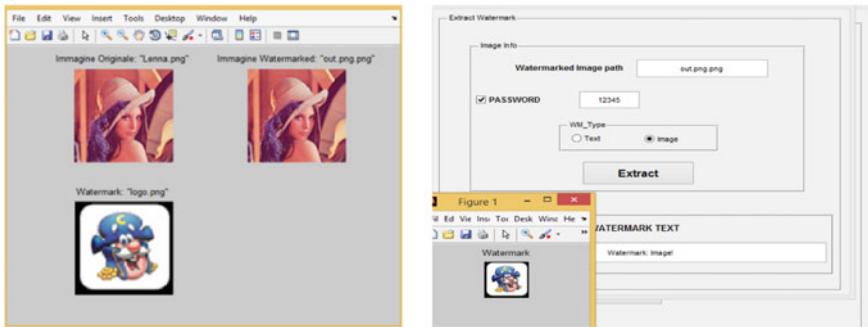


Fig. 5 Original image, watermarked image, extracted watermark

Table 1 PSNR values and NCC of all watermarked images

Attacks	PSNR	NCC
Gaussian noise	45.46	0.9463
Salt and pepper noise	42.38	0.8338
Speckle noise	39.17	0.7842
Median filter	42.63	0.8531

the various type of attacks. Figure 5 shows that the author successfully extracted watermarked image from the original image. The author could verify the integrity [16–19] of the image. Figure 6a, b demonstrates the robustness [20–22] against JPEG compression, Gaussian noise, brightness, and contrast.

In Table 1, the author has calculated the PSNR and NCC values for the color image of Lena 512×512 bit, and the author has taken 128×128 bit watermark color logo image. Then, he inserted the watermark image into the original image, and then, the author has applied various types of attacks onto the watermarked image. After applying the attacks, he has successfully extracted the watermark.

5 Conclusion

In this paper, the author combined both LSB and DWT technique which has been implemented to form a secure and robust digital image watermarking. In the detail coefficient blockade of DWT block, select coefficient set of SVD domains that are inserted into the watermark. Because the author has to increase imperceptibility of the cover image, the cover image can be adjusted by using weighted domain in the frequencies of special area. If the watermarked image attacked by trimming, compression, brightness adjusting, filtering, and noises which can be modified to get the private intellectual property, the process will become more visible and robust. These outcomes conceal that the proposed method is more suitable.

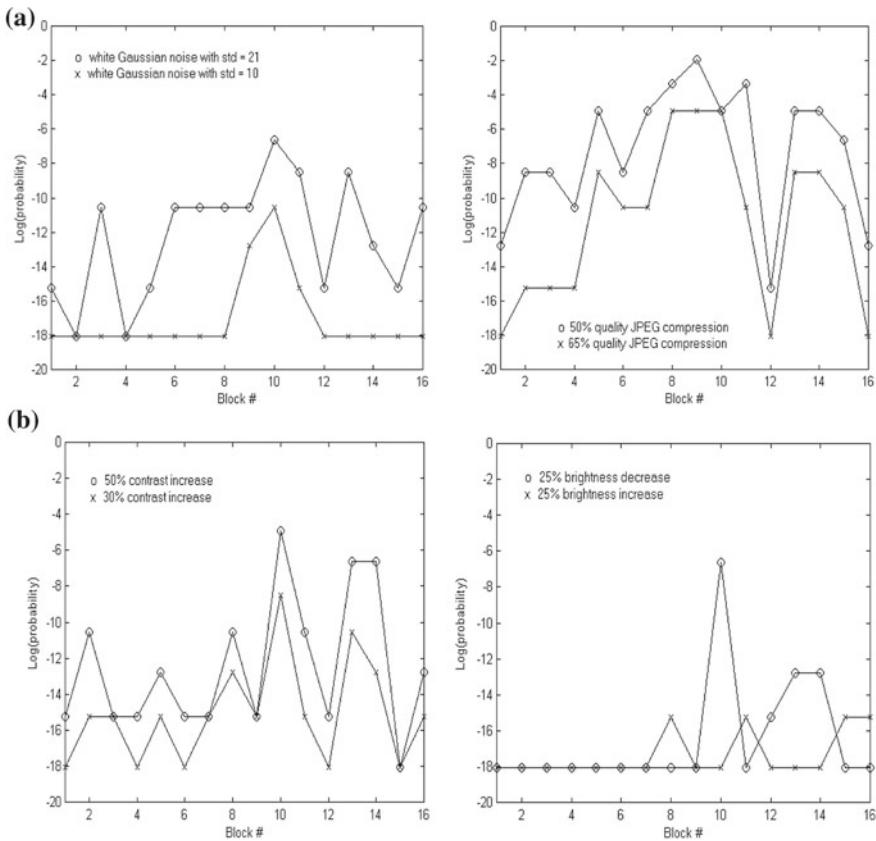


Fig. 6 **a** Robustness against Gaussian noise and JPEG compression. **b** Robustness against brightness and contrast

References

1. Bittman, J., Steinmetz, A.: Content-based digital signature for motion pictures authentication and content-fragile watermarking. In: IEEE proceeding international conference on multimedia computing systems (1999)
2. Lin C., et al.: Robust image authentication method surviving jpeg lossy compression. In: Proceeding of SPIE international conference on storage and retrieval of image/video database
3. Lu, C., Liao, H.: Structural digital signature for image authentication. In: Proceeding of the multimedia security workshop ACM international conference on multimedia
4. Tewari, T.K., Saxena, V.: An improved and robust DCT based on digital image watermarking scheme. Int. J. Comput. Appl. (0975–8887)
5. Lagzian, S., Soryani, M., Fathy, M.: A new robust watermarking scheme based on RDWT-SVD. Int. J. Intell. Inf. Process. **2**(1), 22–29 (2011)
6. Hague, S.R.: Singular value decomposition and discrete cosine transform based image watermarking. Department of Interaction and System Design
7. Rastegar, S., Dr. Bruce.: Watermarking algorithm based on singular value decomposition and radon transform. Int. J. Commun. **65**(7), 658–663 (2011)

8. Kalra, G.S., Talwar, R., Sadawarti, H.: Robust blind digital image watermarking using DWT and dual encryption technique. In: International conference on computational
9. Kumar, A., Tyagi, V., Ansari, D., Kumar, K.: A practical buyer-seller watermarking protocol based on discrete wavelet transform, *Int. J. Comput. Appl.* **21**(8), 46–51
10. Kumar, A., Ansari, D., Ali, J., Kumar, K.: A new buyer-seller watermarking protocol with discrete cosine transform. In: International conference on advances in communication, network, and computing, pp. 468–471. Springer, Berlin Heidelberg
11. Kumar, A., Ghrera, S.P., Tyagi, V.: Modified buyer seller watermarking protocol based on discrete wavelet transform and principal component analysis. *Indian J. Sci. Technol.* **8**(35), 1–9
12. Kumar, A., Ghrera, S.P., Tyagi, V.: Implementation of wavelet based modified buyer-seller watermarking protocol. *WSEAS Trans. signal process.* **10**, 212–220
13. Kumar, A., Ghrera, S.P., Tyagi, V.: A new and efficient buyer-seller digital watermarking protocol using identity based technique for copyright protection, In: Third International conference on image information processing (ICIIP), IEEE
14. Kumar, A., Ghrera, S.P., Tyagi, V.: A comparison of buyer-seller watermarking protocol (BSWP) based on discrete cosine transform (DCT) and discrete wavelet transform (DWT). In: Emerging ICT for bridging the future-proceedings of the 49th annual convention of the computer society of india (CSI) vol. 1, pp. 401–408. Springer. (2015)
15. Kumar, A., Ghrera, S.P., Tyagi, V.: An ID-based secure and flexible buyer-seller watermarking protocol for copyright protection, science & technology. *Pertanika J. Sci. Technol.* **25**(1), 57–76 (2017)
16. Fridrich, J.: Image watermarking for tamper detection. Proceedings of ICIP '98, Chicago (Oct 1998)
17. Walton, S.: Information authentication for a slippery new age. *Dr. Dobbs J.* **20**(4), 18–26 (1995)
18. Van Schyndel, R.G., Tirkel, A.Z., Osborne, C.F.: A digital watermark. In: Proceedings of the IEEE international conference on image processing. **2**, pp. 86–90. Austin, Texas, (Nov 1994)
19. Wolfgang, R.B., Delp, E.J.: A watermark for digital images. In: Proceedings of IEEE international conference on image processing, vol. 3, pp. 219–222. (1996)
20. Ensafi P., Tizhoosh, H.R.: Type-2 fuzzy image enhancement. Systems design engineering, pp. 159–166. University of Waterloo (2005)
21. Hartati, S., Harjoko, A., Nickerson, B.G.: Image enhancement using fuzzy hyperbolization and artificial neural network for anomaly detection. *Int. J. Comput. Electr. Auto. Control Inf. Eng.* vol. 3, pp. 2089–2092. (2009)
22. Kumar, A., et. al.: A lightweight buyer-seller watermarking protocol based on time-stamping and composite signal representation. *Int. J. Eng. Technol.* **7**(4.6), 39–41 (2018)

Time Series Forecasting Using Double Exponential Smoothing for Predicting the Major Ambient Air Pollutants



Rajesh Bose, Raktim Kumar Dey, Sandip Roy and Debabrata Sarddar

Abstract Air quality index (AQI) is the potential air quality measure for the city dwellers which can help to determine the ambient quality of their surroundings. AQI looks like “One number—One Color—One message” which conveys to the common people to take necessary steps. Prediction of AQI is always giving one step ahead from the current situation. The main objective of our research was to forecast the value of major ambient air pollutants using exponential smoothing which is a simple method and also include the trend and seasonality in the sensed data. In this manuscript, the value of major ambient air pollutants is not only forecasted using single exponential smoothing but also introduce double exponential smoothing technique which allows trend on data. Furthermore, our research will be implemented in a fog-enabled data deployment technique which can reduce network bandwidth and also save data storage cost of remote cloud platform. These predicted data will help to identify the green route in the future urban city.

Keywords Air quality index (AQI) · Cloud computing · Exponential smoothing · Fog computing · Green vehicle routing · Internet of Things (IoT)

R. Bose · R. K. Dey
Simplex Infrastructures Ltd, Kolkata, India
e-mail: bose.raj00028@gmail.com

R. K. Dey
e-mail: deyraktim@gmail.com

S. Roy (✉)
Brainware Group of Institutions, Kolkata, India
e-mail: sandiproy86@gmail.com

D. Sarddar
University of Kalyani, Kalyani, India
e-mail: dsarddar1@gmail.com

1 Introduction

Urbanization has a significant result in urban meteorology. Nowadays, over the half of population live in urban areas [1]. Due to the rapid urbanization in Asian and African countries, there are huge demands on infrastructure, job, climate, and environment. Impacts of urbanization degrade the quality of the air which also increase the mortality rate [2, 3]. Due to the poor urban planning and adverse road traffic condition, the pollutants mainly ozone (O_3), particulate matter (PM), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2) are badly affected on human health [4]. Many researchers are now working on the diversity of air pollutants-monitoring projects based on Internet of Things (IoT) [5].

Nowadays, IoT with cloud computing offers various data management tasks which are collected and transmitted by low power and cost microcontroller like Arduino and Raspberry Pi. [6]. The study showed that poor air quality affected the health of the communities and also degraded the climate conditions of our surroundings [7]. Therefore, it is an essential task to monitor the air quality in the city area and alert the urban dwellers on a timely basis. Due to this, different air pollution monitoring units were invented and collected the data from the environment. But now, it is also essential for the green vehicle routing problem because the effective traffic management strategies also improve the traffic-related air pollution in city areas [8, 9]. Researchers suggested that forecasting ambient air pollutants data will also foretell the weather conditions especially at the traffic junctions which will also help to design the green traffic route.

Roy et al. suggested that forecasting atmospheric pollutants data will help to determine the green vehicle route in the urban city areas [10]. The objective of their research was mainly focused on forecasting atmospheric air pollutants using single exponential smoothing but it does not excel on trend data. Due to the frequent change of road traffic congestion in the city areas which impact the quality of air mainly peak hour; so alternative routes do not only improve the traffic conditions but also reduce the air pollutants in the atmosphere.

In this manuscript, our research aim is to forecast the values of atmospheric air pollutants using double exponential smoothing allowing different trending parameter values. Moreover, our research will perform in a fog-based data management scheme which saves network bandwidth and data storage cost at remote cloud platform which has been already discussed in [10].

2 Literature Survey

The related work can be categorized into two different strategies: air pollutant monitoring device using Internet of Things and efficient forecasting model to determine the green vehicle route.

Kadri et al. designed an ambient air quality monitoring unit and forecasting system [11]. Their system wirelessly connected with different distributed monitoring stations and communicated data to the backend server using machine-to-machine communication protocol. They have also implemented an artificial neural network-based prediction system which can foretell the value of air pollutant data before one hour or eight hours.

AIRNOW is a nationwide air quality monitoring program which managed by United States Environmental Protection Agency [12]. It can monitor a specific set of air pollutants which is called criteria air pollutants. AIRNOW offered daily AQI conditions and also predict the AQI values over three hundred cities across the USA.

Choi et al. described a double exponential smoothing model for forecasting CO₂ emissions of the transportation sector in the fifty states and the USA [13]. In [14], the authors explained that how to take preventive measures to protect against ambient air pollution using particulate-filtering facemasks.

In the article [15], it has also noted that the usefulness of measuring the ambient pollutants like nitrogen dioxide (NO₂), carbon monoxide (CO), and sulfur dioxide (SO₂).

Kang et al. highlighted that vehicle emissions are one of the major sources of urban air pollution [16]. Therefore, for roadside air pollution monitoring systems, analyzing and predicting the air pollutant data are required for the urban dwellers to safeguard themselves.

Therefore, our research goal is to predict the ambient air pollutants data like particulate matter 10 (PM₁₀), carbon monoxide (CO), ozone (O₃), oxide of nitrogen (NO_x), nitrogen dioxide (NO₂), and sulfur dioxide (SO₂) at least thirty minutes before time using single and double exponential smoothing techniques [5].

3 Mathematical Formulation and Result Analysis

In our research, we have firstly concentrated on the mathematical formulation for forecasting any time series ambient pollutant data. For our experimental study, we have manually recorded the real-time air pollutant data from Central Pollution Control Board official website on October 17, 2017 for an average of thirty-minute durations in central Kolkata area [10, 17].

To forecast the data, we used the single exponential smoothing (SES) formula for any time period t , which is explained below and documented in [10]:

$$F_t = \alpha C_t + (1 - \alpha) F_{t-1} \quad (1)$$

F_t is called the current smoothed value, C_t is the current observation, and α is called the smoothing constant, where $0 < \alpha < 1$ and $t \geq 3$.

Moreover, the following two equations represent the double exponential smoothing [18]:

$$\begin{aligned} F_t &= \alpha C_t + (1 - \alpha)(F_{t-1} + K_{t-1}) \\ K_t &= \beta(F_t - F_{t-1}) + (1 - \beta)K_{t-1} \end{aligned} \quad (2)$$

The single exponential smoothing can be improved by introducing Eq. (2), whereas K_t represents an estimated trend with a constant γ which is associated with α where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$.

Figures 1a–f and 2a–f and 3a–f illustrate the graphical representation of forecast of major ambient pollutants using single and double exponential smoothing with different $\alpha = 0.1$, $\alpha = 0.5$, and $\alpha = 0.9$ values, $\beta = 0.1$, 0.5 , and 0.9 , respectively.

4 Conclusions and Future Work

Though air pollution is identified as one of the major contributors to the global environmental burden of many diseases, many researchers have already developed different low-cost air quality monitoring devices based on the Internet of Things. In this manuscript, our research goal is to predict the real-time ambient air pollutants data which can alert the urban dwellers to take right precautions. In the future, we will also introduce the trend and seasonality on air pollutants data using triple exponential smoothing technique. Toward the future smart city, our research work will support to identify the green vehicle route, especially for private transport.

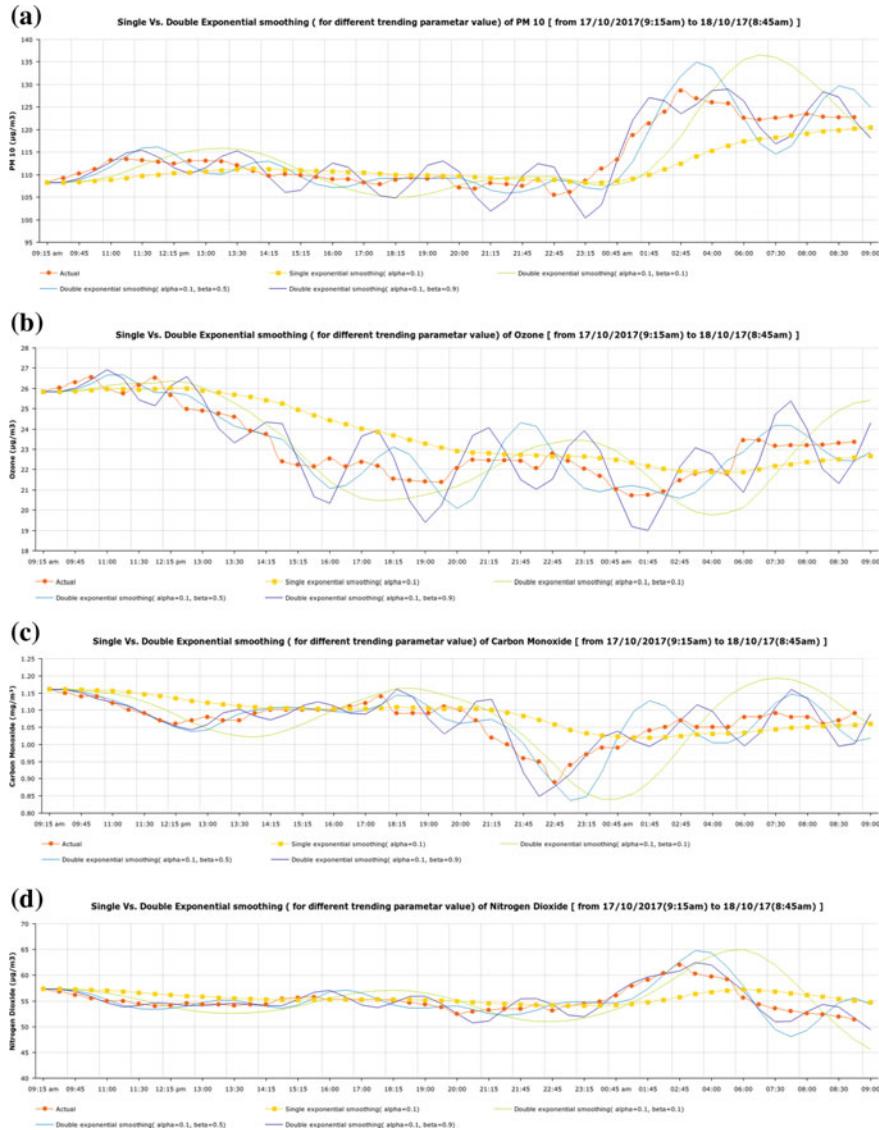


Fig. 1 **a** Forecast with double exponential smoothing on the data of particulate matter 10 from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **b** Forecast with double exponential smoothing on the data of ozone from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **c** Forecast with double exponential smoothing on the data of carbon monoxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **d** Forecast with double exponential smoothing on the data of nitrogen dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **e** Forecast with double exponential smoothing on the data of oxide of nitrogen from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **f** Forecast with double exponential smoothing on the data of sulfur dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.1$ and $\beta = 0.1, 0.5$, and 0.9 , respectively

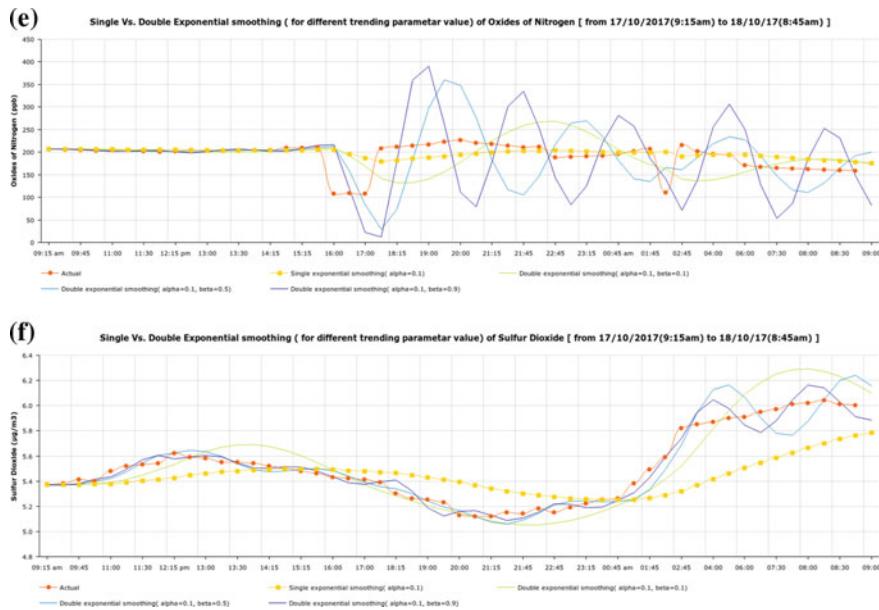


Fig. 1 (continued)

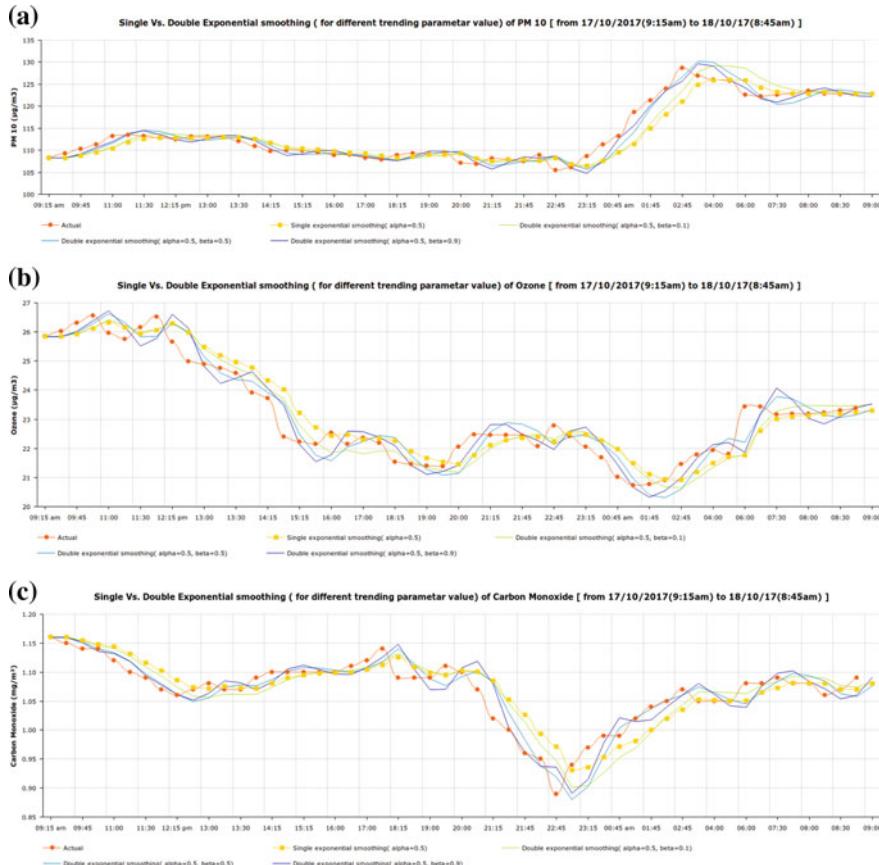


Fig. 2 **a** Forecast with double exponential smoothing on the data of particulate matter 10 from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **b** Forecast with double exponential smoothing on the data of ozone from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **c** Forecast with double exponential smoothing on the data of carbon monoxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **d** Forecast with double exponential smoothing on the data of nitrogen dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **e**, Forecast with double exponential smoothing on the data of oxides of nitrogen from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **f** Forecast with double exponential smoothing on the data of sulfur dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.5$ and $\beta = 0.1, 0.5$, and 0.9 , respectively

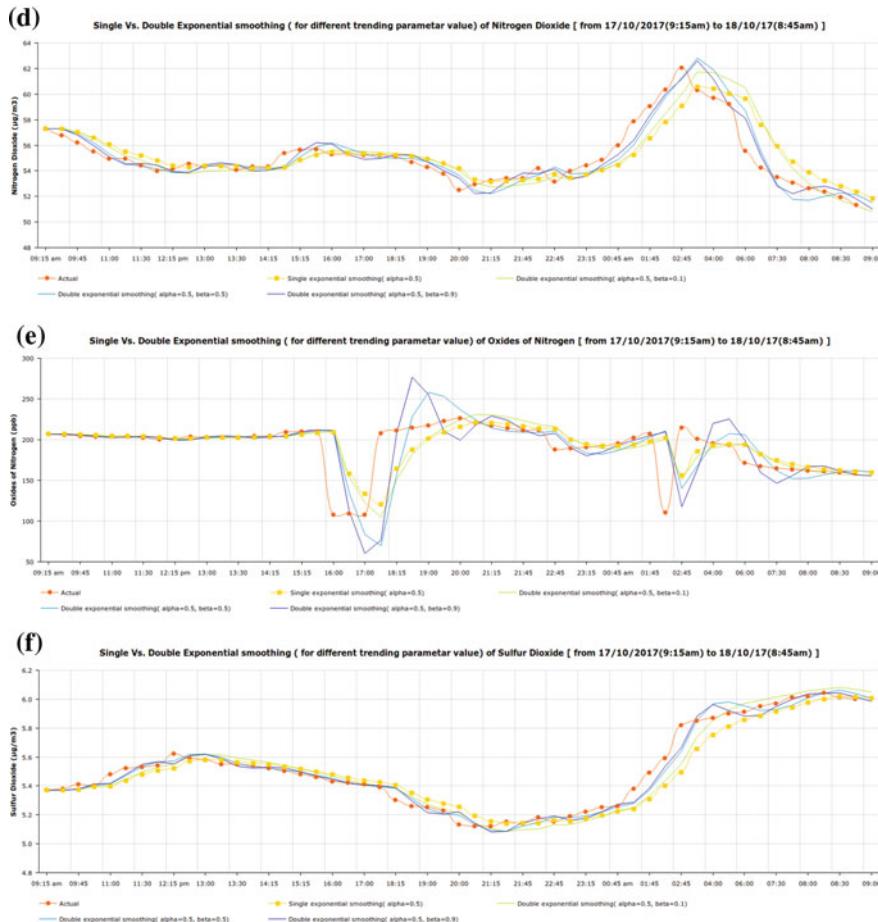


Fig. 2 (continued)



Fig. 3 **a** Forecast with double exponential smoothing on the data of particulate matter 10 from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **b** Forecast with double exponential smoothing on the data of ozone from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **c** Forecast with double exponential smoothing on the data of carbon monoxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **d** Forecast with double exponential smoothing on the data of nitrogen dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **e** Forecast with double exponential smoothing on the data of oxide of nitrogen from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively. **f** Forecast with double exponential smoothing on the data of sulfur dioxide from 17/10/2017 (9:15 am) to 18/10/2017 (8:45 am) with $\alpha = 0.9$ and $\beta = 0.1, 0.5$, and 0.9 , respectively

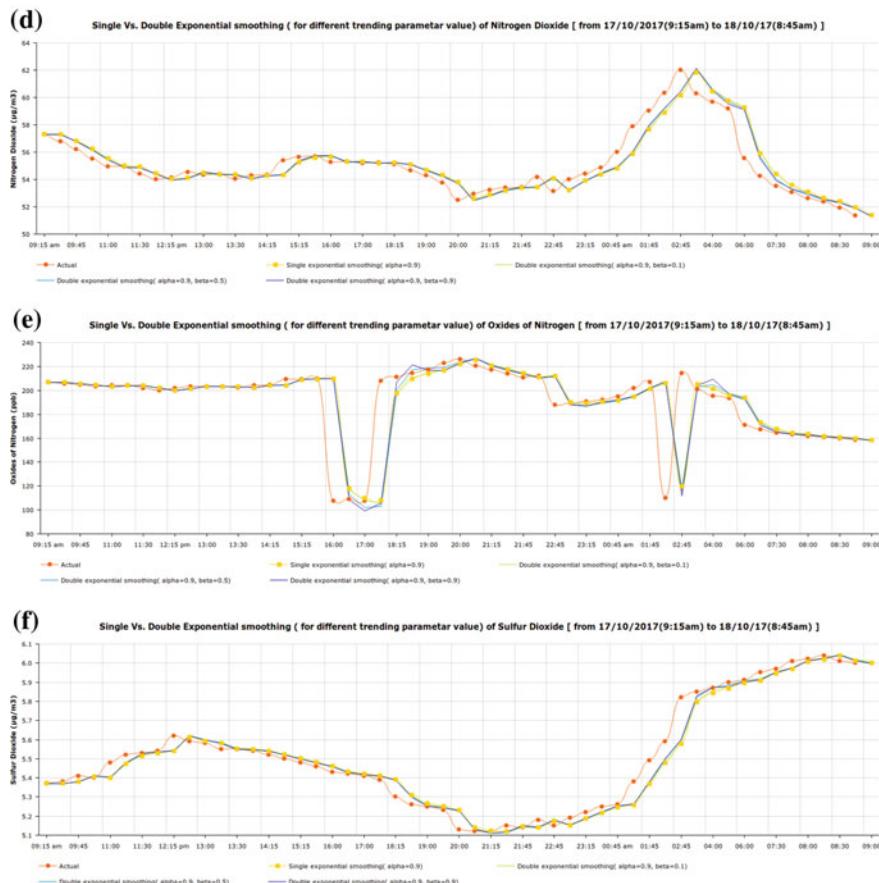


Fig. 3 (continued)

References

1. United Nation.: World Urbanization Prospects: The 2014 Revision. Department of Economic and Social Affairs. pp. 1–32 (2014)
2. Han, L., Zhou, W., Li, W., Li, L.: Impact of urbanization level on urban air quality: a case of fine particles (PM(2.5)) in Chinese cities. *Environ. Pollut.* **194**, 163–170 (2014)
3. Yin, P., He, G., Fan, M., Chiu, K.Y., Liu, C., Xue, A., Liu, T., Pan, Y., Mu, Q., Zhou, M.: Particulate air pollution and mortality in 38 of China's largest cities: time series analysis. *BMJ* **356**(j667), 1–12 (2017)
4. WHO.: Ambient (outdoor) air quality and health. Fact sheets. 1 (2018)
5. Roy, S., Bose, R., Sarddar, D.: Smart and healthy city protecting from carcinogenic pollutants. *Int. J. Appl. Environ. Sci.* **12**, 1661–1692 (2017)
6. Kumar, S., Jasuja, A.: Air quality monitoring system based on IoT using Raspberry Pi. In: International conference on computation, communication and automation (CCCA), Greater Noida, India (2017)
7. Singh, R.P., Prasad, A.K., Chauhan, S.S., Singh, S.: Impact of growing urbanization and air pollution on the regional climate over India. *IAUC Newsletter. Int. Asso. Urban Climate.* **14**, 5–11 (2005)
8. Erdogan, S., Miller-Hooks, E.: A green vehicle routing problem. *Trans. Res. Part E: Logistics Trans. Rev.* **48**, 100–114 (2012)
9. Bigazzi, A.Y., Rouleau, M.: Can traffic management strategies improve urban air quality? A review of the evidence. *J. Trans. Health* **7**, 111–124 (2017)
10. Roy, S., Biswas, S.K., Mahata, S., Bose, R.: Time series forecasting using exponential smoothing to predict the major atmospheric pollutants. In: International conference on advance in computation, communication control and networking (ICACCCN). pp. 1–6 (2018)
11. Kadri, A., Shaban, K.B., Yaacoub, E., Abu-Dayya, A.: Air quality monitoring and prediction system using machine—to—machine platform. In: International Conference on Neural Information. Proceedings on ICONIP 2012. Lecture Notes in Computer Science vol. 7666, 508–517 (2012)
12. AIRNOW: Local air quality conditions and forecasts. (2012) <http://airnow.gov/>
13. Choi, J., Roberts, D., Lee, E.: Forecast of CO₂ emissions from the U.S. transportation sector: estimation from a double exponential smoothing model. *J. Trans. Res Forum* **53**, 63–81 (2014)
14. Zhang, J., Mu, Q.: Air pollution and defensive expenditures: evidence from particulate—filtering facemasks. *J. Environ Econ. Manage.* 1–20 (2018)
15. Nitrogen Dioxyde (NO₂) in our atmosphere. (2017) <http://aqicn.org/faq/2017-01-10/nitrogen-dioxyde-no2-in-our-atmosphere/>
16. Kang, Y., Ding, Y., Li, Z., Cao, Y., Zhao, Y.: A networked remote sensing system for on-road vehicle emission monitoring. *Sci. China Info. Sci.* **60**, 1–10 (2017)
17. Air Quality Data: CPCB official website: <http://cpcb.nic.in/>. (2017)
18. NIST/SEMATECH e-Handbook of statistical methods. Double Exponential Smoothing. (2012) <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc433.htm>

Compression for DNA Sequences Using Huffman Encoding



Raju Bhukya, Subodh Yadav, Jitendra Kumar Sharma,
Banabari lal Sharma and Arun Kumar

Abstract Over the last two decades, the DNA sequence handling and storing problem has been considered as a big problem for many bioinformatics researchers because genomic databases are increasing drastically. To handle this problem, computational biology plays an important role, such as searching for homology, genome formulation, predicting for a new protein sequence, hereditary control networks, and new creative genomics structure. Available resources are not sufficient for storing and handling large DNA sequences. There are various tools developed by using different algorithms and approach. We have proposed an algorithm to improve the compression ratio which uses the concept of RLE (run length encoding) with a Modified HuffBit algorithm. The results obtained by the proposed method are found to be 20% more accurate when compared to existing algorithms.

Keywords HuffBit compress algorithm · DNA sequences compression · Run length encoding

R. Bhukya (✉) · S. Yadav · J. K. Sharma · B. lal Sharma · A. Kumar
NIT Warangal, Telangana 506004, India
e-mail: raju@nitw.ac.in

S. Yadav
e-mail: subodhy53@gmail.com

J. K. Sharma
e-mail: jitujack10@gmail.com

B. lal Sharma
e-mail: banbarisharma09@gmail.com

A. Kumar
e-mail: arunyc2@gmail.com

1 Introduction

The database of the human genome contains around 4 billion characters. So the size of the database is very large, that is why to store, access, and manipulation are very difficult. Storage cost became very high because resources are insufficient to store such type of dataset because it is increasing at a very high rate. So there are two options for this solution: one increases the capacity of the database which is very costly, and the second option is to reduce the size of the dataset without losing any information. Compression is a solution that is able to overcome these challenges, by reducing the storage size and processing costs, as well as increasing transmission speed [1, 2]. Compression should be lossless because there is no meaning of compression if we do not get decompressed data as it is as the original. So there is a requirement of efficient compression algorithm that compressed the DNA sequences in an efficient way.

DNA stands for deoxyribonucleic acid. It is the chemical molecules that are present in all living things' genetic. DNA chain is made from four types of nucleotide: A, C, G, and T; these stand for adenine, cytosine, guanine, and thymine, respectively. All subunit composed of 2'-deoxyribose and phosphate group. Generally, the structure of the DNA cell is a double helix structure. DNA sequences consist of four symbol: A, C, G, and T. The most famous compression algorithm for DNA is two-bit encoding in which every symbol replace with two-bit binary code. In this way, the size of a dataset of DNA is reduced. But DNA sequences are not using the fixed structure of sequences. So compression of DNA sequences is a very challenging task. Many compression algorithms have been developed in the past decade. We will discuss the previous methods as well as a proposed method of DNA compression in brief.

2 Background and Related Work

DNA compression techniques broadly categorized into lossy and lossless [3]. In lossy compression, the whole data is not recoverable after the decompression process. Such algorithms are not useful for compression of DNA sequences. In lossless compression, the whole data is recoverable. There are several existing algorithms for DNA compression discussed.

Huffman coding [4] uses two bits encoding method for constructing a binary tree. Each base replaces by two bits like A = 00, C = 01, G = 10, and T = 11. The required number of bits is equal to just double of a number of bases in DNA sequence. So there is space for modification toward getting a better result. Another compression algorithm, known as GenCompress [5], works on prefix, approximate matches, as well as complimented palindrome. GenCompress algorithm developed as GenCompress-1 and Gencompress-2. GenCompress-1 algorithm uses hamming distance using substitution operation. GenCompress-2 uses edit distance based on

the insert and delete operation. Modified HuffBit compress Algorithm [6] uses the concept of HuffBit compression algorithm using an extended binary tree. In this, required number of maximum bit is equal to three for a base with the lowest frequency. Those bits replace the bases with their corresponding encoded binary code.

Modified HuffBit compress algorithm is better than the two-bit encoding method and HuffBit compresses algorithm in terms of compression ratio and percentage of compression ratio. The Cfact [7] is similar to Biocompress and uses a two-pass algorithm to search for the longest exact and reverse complement repeats. Huffman coding has higher space complexity in most of the cases, and it does not give the best compression ratio. In the worst case, Huffman coding is not efficient. HuffBit Compress [8] algorithm uses extended binary trees to generate the binary code for each base. The base which has max frequency is always stored in the left child of node and base which has min frequency always stored in the right child of node. The values 0 and 1 are assigned for left and right child, respectively. HuffBit Compression algorithm has two steps. The first includes counting the frequency of each base, and the second step includes constructing the extended binary tree. In best case, the algorithm gets compression ratio of 1.006 bpb. The obtained result is not efficient because it scans the whole file for the frequency of each base before constructing the extended binary tree. DNABIT compress [9] is used to compress both repetitive and nonrepetitive DNA sequences. PatternHunter [10] is able to find homologies between sequences as large as human chromosomes, in mere hours on the desktop.

In this paper, we have tried to develop a new algorithm which achieves a better compression ratio than other existing algorithms. Moreover, the proposed algorithm is simple and easy to implement which gives better compression ratio than Modified HuffBit algorithm.

3 Proposed Method

The anterior algorithm is a better approach to compress the DNA sequences but not best, and there is still a space for amendment. We can modify it by utilizing the RLE (run length encoding) approach for achieving better results. In this section, we will show the approach to develop the incipient algorithm.

3.1 DNA Sequence Download

The National Center for Biotechnology Information (NCBI) advances science and health by providing access to biomedical and genomic information. From NCBI, the DNA sequences are downloaded and are saved in fasta format.

Store DNA bases and its frequency in a sorted manner. For compression, there is a need that most of the repeated bases must be replaced by least number of bits because it gives the best compression ratio. That is why the table must be stored

in sorted order, and the length of bases multiplied by their frequency is used as a parameter for sorting. Read the DNA input file and calculate the frequency of each base.

Create an intermediate file with bases. After counting the frequency of bases, we have to create an intermediate file which contains the bases. Like given sequence is AAAAACCCAAAAGGTAATTG, then an intermediate file will be generated which looks like this A5C3A4G2T1A2T2C1G1. This file will be used to compress DNA sequences. Once compression has been completed, the intermediate file is deleted.

Assign the code to each base. The intermediate file we will assign the code to each base on the basis of Length of bases multiplied by frequency has a greater value assigned with the minimum number of bits. The first base which has maximum value will be assigned with ‘0,’ and second base which has second max value will assign with ‘1,’ third base which has third max value will assign with ‘00,’ and so on until all bases have not assigned with some binary code. This is the best way for assigning the binary code to each base. It will give good results when compared to the other discussed algorithms. The reason behind using the table is that the access time of the map is one, and it stores the data (key-value pair) in sorted order.

3.2 *Compression of the Sequence*

For compressing the DNA sequence, the bases will be replaced by their binary codes in an intermediate file which was already generated. The function of compression will use the intermediate file as well as code table which was created by Assigned code Function. Compression function will replace the base by their respective binary codes in intermediate file and save the file as a compressed file. In this way, compression of DNA sequence is made easy. But, the frequency of bases $[m_1, m_2, m_3, \dots, m_n]$ is different for different sequences. So, for different sequences, the different files will be generated. For decompression of the sequence, some information must be given for identifying them from the compressed sequence. We are using the map to store each base and their generated binary code that will be used at the time of decompression.

Compression Algorithm.

Input: DNA sequence fasta format file
sequence.fasta

Ouput:{Compressed file in binary code
}

1. Src[] = copy of file.
2. Generate intermediate file.
3. For i = 0; i < len; i++
Begin for
4. if(src[i] == '\n' || src[i] == ' ')
5. continue;
6. dest += src[i] //append bases
7. rLen = 1;
8. while(i + 1 < len && src[i] == src[i+1])
9. {
10. rLen++ // increment by one
11. i++// increment by one

12. }
13. sprintf(count, "%d", rLen);
14. for (k = 0; *(count+k); k++, j++)
15. dest += count[k]//append count
16. }
17. dest[j] = '\0'// at last null append
18. Count the frequency of each base.
For each base
19. Count_freq[l]++ //increment
20. Generate the binary code for each base.
21. Code[] = {0,1,00,01,.....}
//Replace the bases with their respective binary code.
22. For each base
Replace base by their code.
23. Save the file as a compressed file.

3.3 Decompression of the Sequence

Decompression is the most important part because without it there is no meaning of compression if we will not get the decompressed file as original DNA sequence file. We used the lossless algorithm for DNA sequence compression. So, we will get the original file of DNA sequences without losing any information. For decompression of compressed file, we have to read the compressed file by function. Then, we replaced the encrypted binary code by their respective bases using map. In this way, we will get the intermediate file. Now intermediate file used by a decompressed function will generate the decompressed file as it is original and save the file as a decompressed file. In this way, we will get the original sequence. The algorithm for decompression is given below.

Decompression Algorithm.

1. Read compressed file and create intermediate file
2. Read intermediate file and create decompressed file
3. ifstream fout("intermediate.txt");
4. ofstream fin("compressedfile.bin");
5. char ch
6. while (fin.eof()!=null) //read file till last character
7. begin:
8. string str
9. fin.get(ch) // read char wise from file
10. str+=ch // append chr to get the code
11. fin.get(ch)
12. str+=ch // append chr to get the code
13. fout<<encoded_map[str] //replace binary code by their bases in the file
14. str.clear();
15. End:
16. Save the file as Decompressed file

3.4 Formula and Measurement of Compression Ratio

Compression ratio is represented as a number of bits encrypted to the number of bases, where a number of bases are equal to the number of characters in the DNA sequence.

$$\text{Required number of bits} = \sum F(b) * B(b)$$

Where, $B(b)$ = Number of bits in the encrypted code of bases.

$F(b)$ = Frequency of bases in the sequence.

Example:

Given Sequence, AAACCCTGAAGGGGTTAAGGGAGAAGCAATAG-GAAAA Number of Bases = 38, now function will calculate the frequency for each continuous length of bases as shown in Table 1.

Now store the frequency and bases in a sorted manner using the (length of bases * frequency) as a parameter in Table 2.

To make compression as efficient as possible, we have to assign the minimum number of bits to bases which have a maximum value of (Length of bases * frequency) in descending order. The developed function will assign the code as shown in Table 3. Now we calculated the required number of bits for each base and a total number of required bits as shown in Table 4. where

$$\text{Req_bits} = \text{Freq} * \text{No_bits}$$

Compression Ratio = Total Number of Required bits/number of base Ratio,

$$\text{Compression Ratio} = 40/38 = 1.05263$$

Table 1 Store the frequency and their bases

Base	Frequency
A3	1
C3	1
T1	2
A2	4
G1	3
G4	1
T2	1
G3	1
A1	2
C1	1
A5	1

Table 2 Frequency and bases in sorted order

Base	Len	Freq	Len*freq
A2	2	4	2*4 = 8
A5	5	1	1*5 = 5
G4	4	1	1*4 = 4
A3	3	1	1*3 = 3
C3	3	1	1*3 = 3
G3	3	1	1*3 = 3
G1	1	3	3*1 = 3
T2	2	1	1*2 = 2
A1	1	2	2*1 = 2
T1	1	2	2*1 = 2
C1	1	1	1*1 = 1

Table 3 The binary code assigned to each base

Base	Len*freq	Assigned code	No_bits
A2	4*2 = 8	0	1
A5	1*5 = 5	1	1
G4	1*4 = 4	00	2
A3	1*3 = 3	01	2
C3	1*3 = 3	10	2
G3	1*3 = 3	11	2
G1	3*1 = 3	000	3
T2	1*2 = 2	001	3
A1	2*1 = 2	010	3
T1	2*1 = 2	100	3
C1	1*1 = 1	110	3

Table 4 Store the number of required bits

Base	Freq*Len	No_Bits	Req_bits
A2	$4*2 = 8$	1	$4*1 = 4$
A5	$1*5 = 5$	1	$1*1 = 1$
G4	$1*4 = 4$	2	$1*2 = 2$
A3	$1*3 = 3$	2	$1*2 = 2$
C3	$1*3 = 3$	2	$1*2 = 2$
G3	$1*3 = 3$	2	$1*2 = 2$
G1	$3*1 = 3$	3	$3*3 = 9$
T2	$1*2 = 2$	3	$1*3 = 3$
A1	$2*1 = 2$	3	$2*3 = 6$
T1	$2*1 = 2$	3	$2*3 = 6$
C1	$1*1 = 1$	3	$1*3 = 3$
Total			40

4 Results

The proposed algorithm is used for better DNA compression compared with existing algorithms. It is an enhanced version of the modified HuffBit compression algorithm using RLE (Run Length Encoded). The existing algorithms are good for DNA compression but there is a lot of scope for improvement. The proposed algorithm is better from the existing algorithms on the basis of the following parameters.

- (i) Compression ratio.
- (ii) Percentage of compression ratio.

We will compare the proposed algorithm with the previous algorithms.

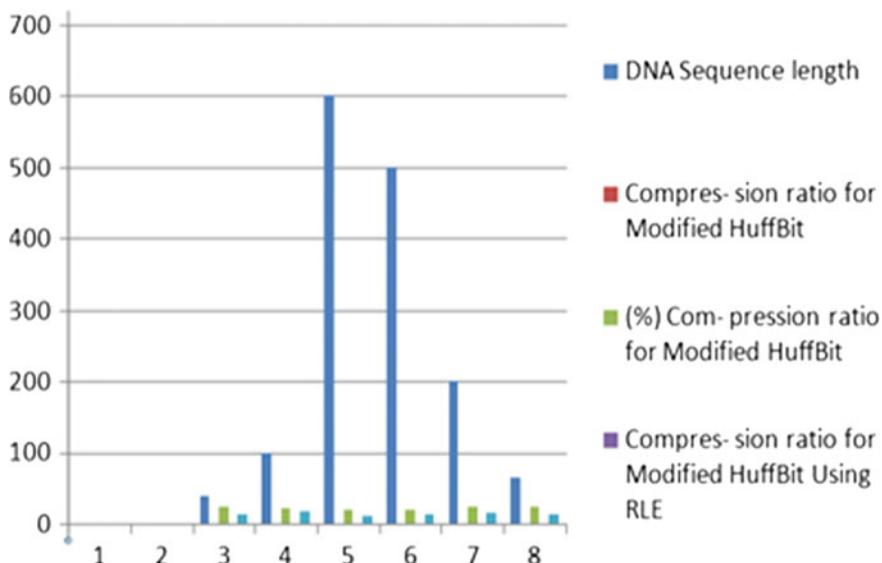
4.1 Comparison with a Prior Algorithm

There are many compression algorithms developed by researchers, and improvement is going on for DNA sequences compression. The proposed algorithm gives better results in comparison with other existing algorithms. The comparison is done on the basis of compression ratio, the percentage of compression ratio, as well as time complexity. The compression ratio can be calculated as the ratio of the size of the uncompressed file to the size of the compressed file. Comparison results between modified HuffBit algorithm and modified HuffBit using RLE (run length encoding) algorithms are tabulated in Table 5. The comparison results are shown in Fig. 1.

Table 5 Comparison of the proposed algorithm and the existing one

DNA Seq. length	Compression ratio for modified HuffBit (bpb)	(%) Compression the ratio for Modifie_d HuffBit compress	Compression ratio for modified HuffBit using RLE (bpb)	(%) Compression ratio for modified HuffBit using RLE
38	1.89	23.68	1.05263	13.15
100	1.71	21.37	1.0325	17.50
600	1.635	20.31	1.0278	11.82
500	1.612	20.15	1.04351	13.58
200	1.98	24.75	1.0670	14.31
64	1.625	24.25	1.1230	12.39

Percentage Compression ratio = (Number of byte encoded/No.of bases) × 100%

**Fig. 1** Compression ratios to corresponding input size

5 Conclusion

Compression of DNA sequence is a very important topic for minimizing the cost of storage in the database. The Proposed Algorithm compresses the large file of DNA sequence in a very easy and efficient way. The main advantage of the proposed algorithm is that it gives the best compression ratio on large data set of DNA sequence. In this way, a large amount of space in the database is saved. Moreover, it takes less time, less space, and is easier to implement.

References

1. Kahn, S.: On the future of genomic data. *Science* **2011**, 728–729 (2011)
2. Alberti, C.: Investigation on genomic information compression and storage.ISO/IEC JTC 1/SC 29/WG 11 N15346; ISO: Geneva, Switzerland, pp. 1–28.(2015)
3. Bharti, R.K., Harbola, D.: State of the art: DNA compression algorithms. *IJARCSSE* 2013; p. 397 (2013)
4. Bakr, N.S., Sharawi, A.A.: DNA lossless compression algorithms: review. *Am. J. Bioinform. Res.* 2013 pp. 72–81 (2013)
5. Chen, X., Kwong, S., Li, M.: A compression algorithm for DNA sequences. *IEEE Engineering in Medicine and Biology Magazine*, pp. 61–66 (2001)
6. Habib, N., Ahmed, K.: Modified HuffBit compress algorithm—an application of R. *J. Integr. Bioinf.* 1–13 (2018)
7. Rivals, E., et al.: Fast discerning repeats in DNA sequences with a compression algorithm. In: The 8th workshop on genome and informatics, (GIW97), 1997, pp. 215–26. (1997)
8. Rajeswari, P.R., Apparao, A., Kumar, R.K.: HUFFBIT COMPRESS—algorithm to compress DNA sequences using extended binary trees. *J. Theor. Appl. Inf. Technol.* pp. 101–106 (2010)
9. Rajarajeswari, P., Apparao, A.: DNABIT compress—genome compression algorithm. *Bioinformation* **2011**, 350–360 (2011)
10. Ma, B., Tromp, J., Li, M.: PatternHunter—faster and more sensitive homology search. *Bioinformatics* 2002 pp. 440–445 (2002)

Use of Clustering Sentiments for Opinion Mining: An Experimental Analysis



D. Veeraiah and J. Nageswara Rao

Abstract In E-commerce, the social media network analytics are in a key position for the extraction of a service or product information. Opinion mining is now the key ingredient for social network analytics. In the current study, we study opinion mining process in networks while different kinds of documents are dealt with that are opinionated and a formal discussion on the challenges for sentiment analysis via social networks is being updated [1]. The social network Twitter has now become a huge online platform which has millions of people with their opinions shared as a social activity [2]. One major concern is how to get reviews for products on the basis of its features. The application of the K-means algorithm on a dataset sample from twitter and twitter dataset which will be clustered by different opinions in the context of various product features and has been evaluated along with explanation by the hand of a machine learning tool.

Keywords Sentiment analytics · K-means algorithm · Twitter · Opinions

1 Introduction

Currently, trending social networks have many features or services like the discussions, reviews, and chats; theme for social media networks is building a social relationship with one another in a society. The tool used to understand and interpret the social media data is called the social media analytics. In view of the rapid growth rate in the social media activities on the web, every individual or organization is considering the previous opinions and reviews for the decision making. With all these diverse effects, there has been a lot of data and is very difficult for anyone for identifying the corresponding sites and encapsulation of data in them. Figure 1 shows

D. Veeraiah (✉) · J. Nageswara Rao
Dept. of CSE, Lakireddy Bali Reddy College of Engineering, Mylavaram, A.P, India
e-mail: veeraiahdvc@gmail.com

J. Nageswara Rao
e-mail: nagsmit@gmail.com

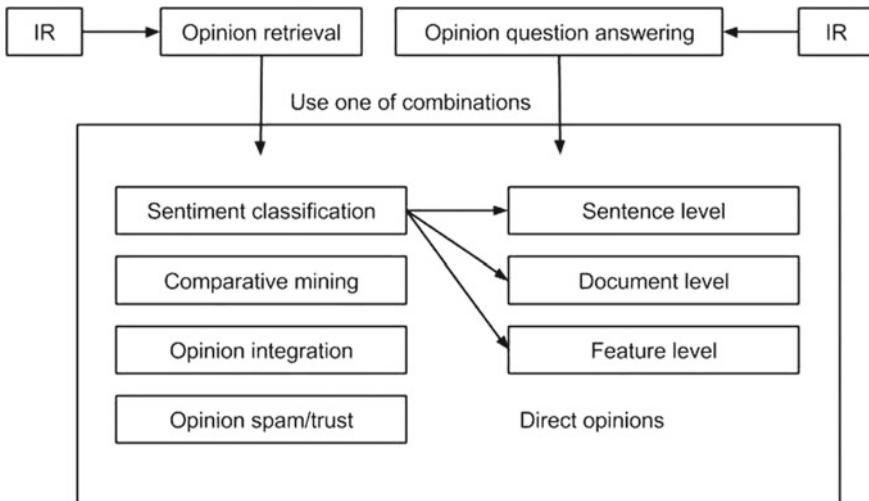


Fig. 1 Opinion mining in social networks

the sentiment analytics in social networks outlay. From information retrieval (IR) actions, information is gathered whose source of information will be some online sources and feedback of the surveys are processed and used for sentiment mining

2 Sentiment Classification

As the reviews or opinions are mixed up feelings and sentiments, the classification has to be done which is called as the sentiment classification. This deals with the sentiment existing in the sentence or feature expressed by the users that can also be +ve or -ve or sometimes neutral. The classification is as follows

- Sentence level: This species whether a sentence or given text transcript is +ve or -ve or sometimes neutral.

Ex: India won the masters championship—2018.

Document level: This does not consider a single statement but considers an entire document for the expression classification whether +ve or -ve.

Ex: Review and polls of products online.

Feature level: Without looking out directly at the language phrases as review and relationships (linguistic). It is on the basis of an idea that is what a review or sentiment comprises of? A +ve or -ve or sometimes neutral sentiments.

- Comparative mining: This is concerned with the extraction of comparative relations and sentences.

Ex: “She is fit”—directly said opinion

“Working is better than wasting time”—comparatively said sentence

Table 1 Opinions on twitter features by students

Question	Good	Bad	Very bad	Very satisfied
Business	48	6	3	0
Education	35	16	2	4
User friendly	19	23	10	5
Logo	37	5	5	10
Business services	24	8	8	17
Interests	23	9	2	23
Groups	15	10	10	22
Advertising	43	4	6	4
Design	42	7	7	1
Notifications	29	16	3	8
Game apps	28	15	9	5
A/c dress Up	28	15	9	5
Shopping	23	15	14	5
Knowledge	35	5	5	12
Useful	17	20	15	5

Integrating of opinion is to integrate the opinions from variable sources present. In view of the rapid growth rate in the social media activities on the web, every individual or organization is considering the previous opinions and reviews for the decision making. Opinion spam is a major issue faced by the current opinion mining research [3]. Opinion retrieval is the process of retrieving of the documents and files depending on the opinions on that specific context and is shown in Table 1.

The graph represented for the above opinion data is given as shown in Figs. 2 and 3.

3 Challenges

Concerns of social networks are with the user behavior and his passions are the wealth of information. This is tough as the types of languages and text forms change from place to place which can be called as multi-languages. The following data table provides the data of both the Greek and English sentences. The classification of the sentiment between the European Union and British is shown in below Table 2.

Fig. 2 User feedback for Twitter

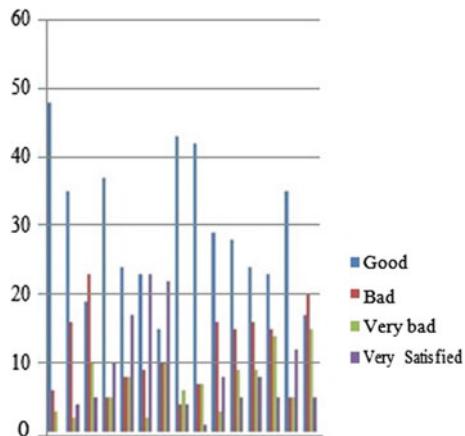
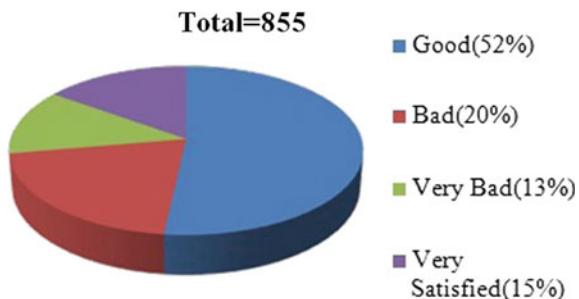


Fig. 3 Percentage levels of user feedback for Twitter



3.1 K-Means Clustering

As per Internet-based life and its client examination, gathering of comparable obstinate individuals is called as grouping. Bunching alludes to a technique by which the datasets of Web clients are assembled into groups of littler sets with comparable information. One of such calculation utilized for bunching is k-implies used to partition the datasets into ‘k’ groups where k is a positive whole number [4].

Algorithm

1. Randomly select c centroids.
2. Calculate distance from each centroid vz to all data points xi .
3. Repeat
4. Assign each data point xi to the centroid with minimum distance.
5. Recalculate new centroids from each new.
6. Recalculate distance from each new centroid to all data points.
7. Until no data point was reassigned a new centroid.

Table 2 Anglo-European dataset sentiment analysis [5]

What the British say	What the British mean	What others understand
I hear what you say	I disagree and do not want to discuss it further	He accepts my point of view
With the greatest respect	I think you are an idiot	He is listening to me
That's not bad	That's good	That's poor
That is very brave proposal	You are insane	He thinks I have courage
Quite good	A bit disappointing	Quite good
I would suggest	Do it or be prepared to justify yourself	Think about the idea, but do what you like
Oh, incidentally/by the way	The primary purpose of our discussion is.	That is not very important
I was a bit disappointed that	I am annoyed that	It doesn't really matter
Very interesting	That is clearly nonsense	They are impressed
I'll bear it in mind	I've forgotten it already	They will probably do it
I'm sure it's my fault	It's your fault	Why do they think it was their fault?
You must come for dinner	It's not an invitation, I'm just being polite	I will get an invitation soon
I almost agree	I don't agree at all	He's not far from agreement
I only have few minor comments	Please rewrite completely	He has found a few typos
Could we consider some other options	I don't like your idea	They have not yet decided

4 Results

Figure 4 represents various clustering opinions of various products & the numerical values are discussed previous. The final output of clustering techniques of different product opinions are shown Fig. 5. The % levels of getting positive views of sample data is shown in Table 3. Between them, 13.5 rates for Apple, 15% for Google, four rate for microsoft and 7% for Twitter are founded and grouped by using k-means technique by collecting the +Ve opinions and reviews of the clients, we find group of users with help of twitter id's from the various data set. So that is help to an organization to spreading its services to various target audience.

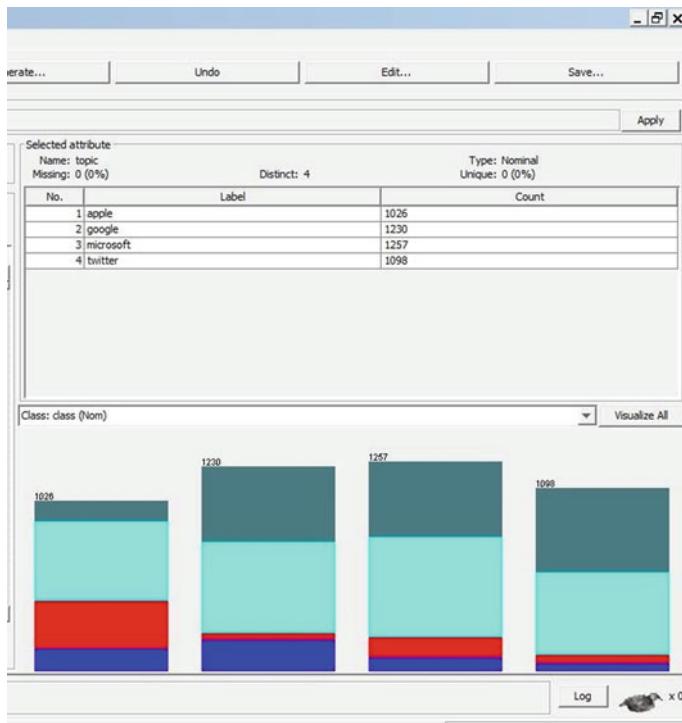


Fig. 4 Clustering of products by opinions

Table 3 Target audience opinions on product/service

Product	Positive (Cluster0)	Negative (Cluster1)	Neutral (Cluster2)	Irrelevant (Cluster3)
Apple	139	294	478	127
Google	187	44	549	460
Microsoft	52	66	494	329
Twitter	77	99	600	616

5 Conclusion

This paper activity is the way of supposition mining the use of content investigation and discussed about to sum things up about the awful circumstances in sentiment mining. It is masses needed for a venture to understand the evaluating of clients about its items. Nostalgic assessment encourages in assessment of valuable substance inside the content. Affirm way is done to the twitter dataset by methods for utilizing part the information into momentous classes of bunches taking feeling as a thought detail. The similar appraisal of various stocks and one among type assessments has been

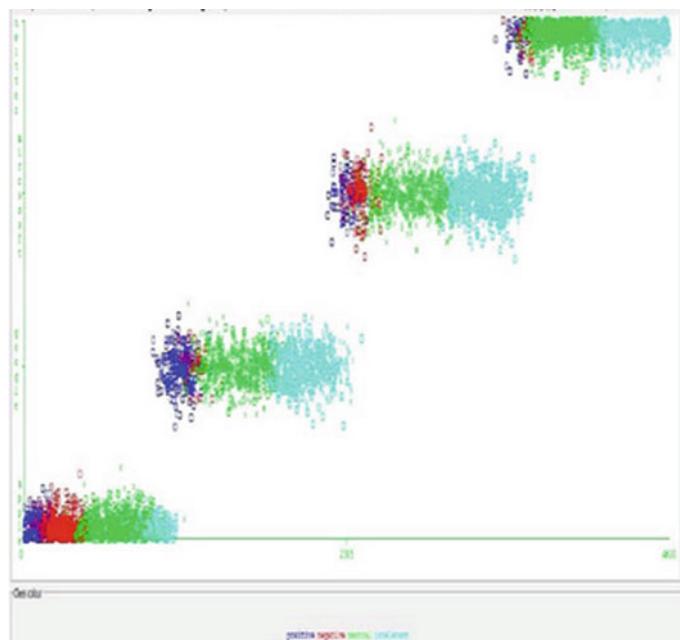


Fig. 5 Application of clustering on product versus sentiments

portrayed. Along these lines, the group of clients who have an enormous sentiment on the item is mulled over and completed alright way bunching set of rules. Basically, so it influences a smooth way to approach the brilliant focus to advertise right now who can imitate the development and amazing of an office. This paper discusses the supposition mining of clients perusing printed content records. The work might be additional material gaining practical experience in heterogeneous insights, for example, photographs, movement pictures for assessment mining. Most extreme of the opinion assessment examine is doing at the capacities top notch and poor groups; anyway in this paper, more bunches unbiased and close to the factor are presented. This makes the experience of oddity and commitment.

References

1. Fan, W., Gordon, M.D.: Unveiling the power of social media analytics, University of Michigan, Ann Arbor, Forthcoming at Communications of the ACM
2. Joshi, N.S., Itkat, S.A.: A survey on feature level sentiment analysis, Int. J. Comput. Sci. Inf. Technol. (IJCSIT), 5(4), 5422–5425 (2014)
3. Maynard, D., Bonccheva, K., Rout, D.: Challenges in developing opinion mining tools for social media. In: Proceedings of @NLP can u tag#usergeneratedcontent?! Workshop at LREC 2012, Istanbul, Turkey (May 2012)

4. Bhatia, M.P.S., Khurana, D.: Experimental study of data clustering using k-means and modified algorithms. IJDKP, vol. 3, No. 3, (May 2013)
5. <http://www.oxfamblogs.org/fp2p/wp-content/uploads/Anglo-vs-EU.jpg>

ICT for HR in an Industrial Organisation for a Small but High-Impact Process (a Case Study of RINL, Visakhapatnam Steel Plant, India)



K. Satya Narayana, G. V. Ramesh and Debasish Ray

Abstract In India, national economy is largely depended on the production process. The industrial work force as a group is one of the prime stake holders over here. Apart from the basic wages, providing additional welfare facilities has become necessary not only as a motivator for improved productivity and production, but also as a moral obligation on the part of the employer. Health concerns of their employees are one of the important areas to be addressed by the organisations as part of the welfare measures. Many organisations provide health facilities through Mediclaim insurance to its employees, wherein cashless treatment is available up to a certain extent. In case of RINL, one of the major public sector steel organisations in India, whose case study is being discussed in this paper, provides primary medical facilities through its in-house hospital. However, for the super speciality medical aid, employees are referred to super speciality hospitals on case-to-case basis through a standardised “medical referral process” devised for this purpose. The authors made an attempt to showcase their study of this case to explain the unforeseen advantages with the use of information and communication technology (ICT) for a simple process, and how it could yield in terms of tangible benefits as well as the intangible upkeep of morale of employees.

Keywords HR · ICT · Employees’ welfare · Steel industry · Medical referral

K. Satya Narayana (✉) · D. Ray

Human Resource Department, Rashtriya Ispat Nigam Limited, Visakhapatnam, AP, India

e-mail: Satyam.dc@vizagsteel.com

URL: <https://www.vizagsteel.com>

D. Ray

e-mail: debasishray@vizagsteel.com

URL: <https://www.vizagsteel.com>

G. V. Ramesh

Information Technology Department, Rashtriya Ispat Nigam Limited, Visakhapatnam, AP, India

e-mail: gvr@vizagsteel.com

URL: <https://www.vizagsteel.com>

1 Introduction

Industrial peace can be ensured through labour welfare and job satisfaction, which are very important dimensions for ensuring harmonious industrial relations and progress of the nation. There are various theories and diverse approaches, through which the organisations provide labour welfare facilities. Medical and educational services, a congenial work environment, etc., are provided to the employees as a social responsibility of industry for upholding democratic values and concern for employees [1]. In India, the provision of statutory welfare measures is a necessary prerequisite for any industrial organisation. Some organisations go beyond the statutory welfare measures and provide additional measures, called non-statutory amenities.

The welfare benefits not only raise employee morale, but also make it easier for employers to attract and hire more competent personnel. Welfare also helps in building a positive image of the organisation and facilitates easier dealing with the workers' unions [2].

It is evident from the research that human resource department plays a major role in the functioning of the organisation by coordinating the work of other departments for the welfare of human resources. It is perceived that the use of ICT in HR practices and processes resulted in cost and time-effective benefits while improving the satisfaction/delight of the workforce. Historically, lack of access to ICT was defined as "digital divide", a separation of society, based on the fact that who have and who don't have effective access to digital technology. The use of ICT in industrial scenario goes a long way in bridging this digital divide [3].

One of the most important concerns of any employee, for that matter any person, is the health—his and his family members. RINL has a well-equipped hospital to take care of most of the generic diseases and health issues. However, the requirement of super speciality facilities and expert consultants is a necessity in some of the cases, which are, presently, not available in the in-house hospital. In order to obviate it, the organisation follows a process called "medical referral process" to provide cashless treatment to the employees and their dependents in some select super speciality hospitals—which the organisation entered into MOU from time to time—across the country.

Information and communication technology (ICT) is widely used in the production process in RINL. Information technology in the form of SAP-ERP is also deployed in most of the HR processes. There are many areas which directly impact the morale of the employees, wherein IT could be further used to simplify the process for providing better facility. It has been identified that medical referral process is one such process, which if brought under ICT gamut, would have greater impact on employees' welfare. Recently, using ICT, medical referral process is reengineered, simplified, made online with user-friendly interface which resulted in many unforeseen advantages to the organisation and to the employees. As a case study, the same process is discussed in this paper.

2 Profile of RINL/VSP

2.1 Capacity and International Standards

Rashtriya Ispat Nigam Limited (RINL)—a Navratna Public Sector Enterprise with 100% ownership of GoI—is the corporate entity of Visakhapatnam Steel Plant. VSP is India’s first shore-based integrated steel plant, located at Visakhapatnam, and has production capacity of 7.3 million tonnes of facilities.

A pioneer in the steel industry in the adoption of system standards, VSP, is accredited for ISO 9001, ISO 14001, OHSAS 18001, CMMI Level 3 for IT system and ISO 27001 for information security management system and is the first PSE to be certified for ISO 50001 for energy management system.

2.2 Technological Facilities, Employee Profile, and Accolades [4]

The process of steelmaking at VSP is through the BF-BOF route. Technologies adopted during inception of plant were the first of its kind in the Indian steel industry, and this has been a source of competitive advantage in operational efficiency. It is further strengthened with stabilization of 6.3 Mt expansion with the latest technological features.

Around 17,600 regular and another 17,000 outsourced employees are engaged in various processes. Eighty-five per cent of the human resource is used in the production process of steelmaking from basic raw materials. As an organisation, over a period of 20 years of growth, RINL achieved many laurels in steel sector from Government as well from other private agencies. Some of them are:

- Star Performer Award 2014–15 and 2015–16 by Engineering Exports Promotion Council, Ministry of Commerce, GoI for Exports Excellence;
- Corporate Vigilance Excellence Award by IPE for commendable work in preventive vigilance;
- “Best Enterprise Award” under Navratna category by SCOPE—first prize for the outstanding contribution for the betterment of women employees;
- Awards at INSSAN—for “Excellence in Suggestion Scheme” and also for excellence in organisational performance;
- “QCFI award for Best Public Sector organisation-2016” for innovation in operational efficiency and contributing towards implementation of Quality Circles;
- “CIO-100 Award” by International Data Group for operational and strategic excellence in Information Technology;
- PSE Excellence Award 2015—second prize for good corporate governance;
- IIIE Performance Excellence Award 2015 for outstanding contribution made towards excellence under “Organisation category”.

3 Use of ICT in RINL/VSP

There has been a wide use of ICT in almost all the manufacturing and business processes for more than two decades with legacy applications. Subsequently, ERP was implemented in supporting business processes like sales and marketing, purchase and inventory management, finance and costing and HR with the integration.

However, compared to other business processes, the implementation of ICT in HR area is not so easy [5]. The reason could be that it has interface with various types of employees with different educational, cultural, attitudinal, and experience backgrounds. In these areas, home-grown IT application systems would be more suitable as their customisation is closely linked with the organisational culture [6]. In this milieu, “medical referral process” is enabled with custom-developed in-house IT system, which has integration with ERP interface.

4 Healthcare at RINL

4.1 *Medical Facilities*

The company provides comprehensive healthcare scheme involving in-patient and out-patient treatment in company run hospitals and specialised treatment at referral hospitals to the employees and their dependants.

Visakhapatnam Steel Plant has established a 160-bedded ultra modern hospital called Visakha Steel General Hospital (VSGH) equipped with sophisticated equipment. In addition to this, there are four health centres at other locations, where employees reside outside the township (including its captive mines' locations), and two emergency medical care units in the plant premises to cater to the health needs of the employees established.

4.2 *Medical Referral Process*

Keeping good health of the employees is one of the most important considerations for HR management in RINL. Since the hospital is run by the steel industry and is mainly meant for employees and their families, medical experts of various faculties are brought from corporate hospitals for consultation. Actual treatment for major diseases, if any identified by medical experts, is carried out in some of the best hospitals across the country and in major cities of Andhra Pradesh. For the purpose of cashless treatment to the employees and their family members at an optimum cost, RINL enters into an MOU with major hospitals across the country. Around three hundred patients get admitted into these hospitals every month, and credit letters are

issued to these hospitals for the treatment, which are subsequently redeemed based on the bills submitted and treatment provided.

4.3 The Legacy Manual Process of Obtaining Credit Letter

- Patient (employee/dependent) gets referred by Visakha Steel General Hospital.
- Consult doctor at referred hospital and get the estimate.
- Request colleague/reporting officer to initiate the note for approval from the competent authority.
- Colleague/reporting officer prepares the note with the help of PAs and sends to HODs of the respective department.
- Then, the physical file is sent to zonal HR personnel for the verification of dependency.
- From zonal HR personnel, after HOD signature, the note sheet is physically taken to hospital and handed over to medical administration department.
- Medical administration collects all such note sheets and prepares a statement for Medical Board.
- Call for Medical Board meeting at 12 AM every day.
- Medical Board clears all straight cases and asks the employee/colleagues for clarifications in other cases.
- After Medical Board clears the proposal, then they are sent for the signature of HOD of Medical.
- After the HOD-medical signature, medical administration prepares credit letters for all cleared cases.
- Employees or colleagues have to come physically and collect the credit letter or give clarification.
- Employee/colleague has to carry the credit letter to the concerned referred hospital and handover it to the hospital authorities.
- After verification by external Hospital authorities, treatment will be started, if all details are found to be in order.
- Generally, the validity of the credit letter is 30 days from the date of the issue of the credit letter. After 30 days, if the treatment is to be continued, this process needs to be repeated (Fig. 1).

It can be seen from the above, the complexity, laboriousness and the number of man-hours put in for each case for obtaining a credit letter. Over a period of time, it was observed that in addition to many intangible disadvantages, there are around 47,800 (@ 12 man-hours for each credit letter) man-hours lost every year for this process.

HR department along with IT Department took up the initiative to simplify the process through reengineering and by developing a suitable Web-based online system.

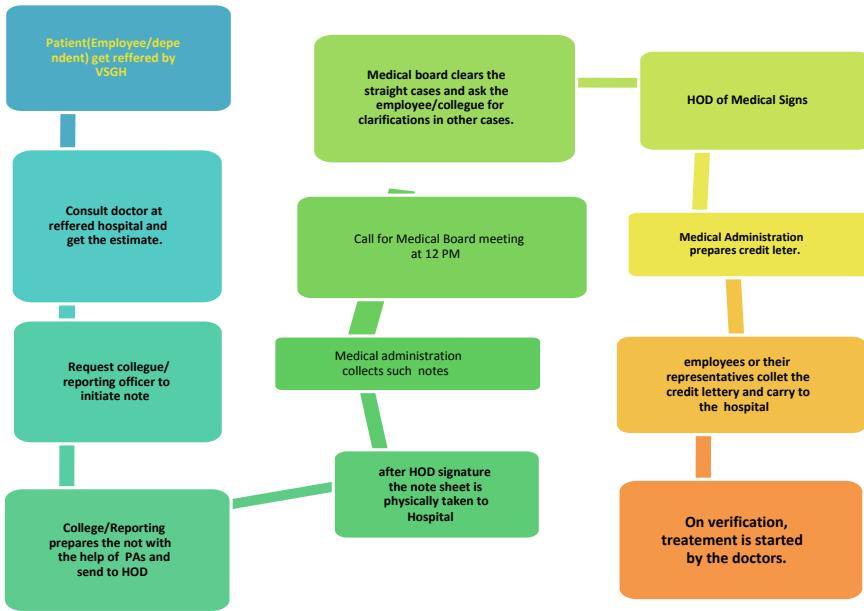


Fig. 1 Manual legacy process flow

4.4 Modified Online Process of Obtaining Credit Letter

- Employee/Colleague of the respective department of the employee logs into the online system from anywhere and selects the Medical ID of the patient.
- Upload a JPG or PDF of the referred hospital estimate (or a zipped file, if it contains more than one file).
- Medical administration periodically allots all the requests received online to the respective Medical Board members which are cleared as and when they are received online.
- As soon as HOD Medical clears the recommendations online, an automated SMS is generated and sent to the employee, HOD of the respective department, medical administration and to the referred hospital.
- Employees or their dependents can directly go to the referred hospitals and get admitted for treatment.
- All the MOU hospitals are given user credentials to access the system through the Internet. They can download the credit Letter and upload the discharge summary (after the treatment is over) (Fig. 2).

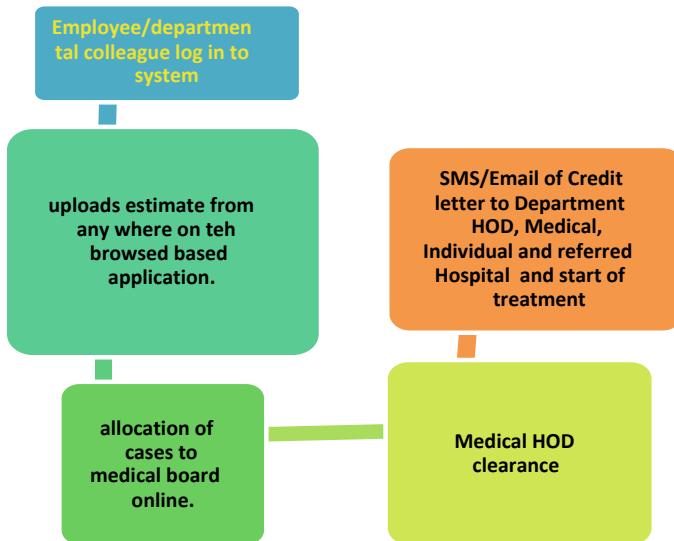


Fig. 2 Modified online process

4.5 Technical Details

- It is a Web-based application using J2EE technology.
- Oracle is the database.
- The employees and medical department personnel use the application deployed in corporate intranet.
- The MOU hospitals use the system deployed in DMZ through the Internet using secure communication protocol (HTTPS).

5 Constraints Faced During the Implementation

In the manual legacy process, there was a manual involvement of colleagues, union leaders, HOD, zonal personnel, medical administration personnel, Medical Board and the referred hospital. Since the proposal pass through all these levels, they feel empowered to clear the proposal, though there is no value addition at most of the stages. However, for the computerisation, major resistance was shown by the union personnel. Union personnel use the opportunity of arranging a credit letter to an employee as a special service to the employees that is being rendered by union. Thus, they attract the employees to get them into their fold. They act like middlemen between employee, HOD, medical administration, and Medical Board. In reality, there was no value addition by them except carrying the note sheet physically to the

HOD office, and to the medical department, and get the credit letter from them, and handover to employee.

The union personnel were convinced through various rounds of negotiations by HR and IT experts by explaining to them the advantages of the automated online medical referral process.

6 Tangible and Intangible Advantages

There are many unanticipated tangible and intangible advantages realised through the online automated medical referral process. Some of them are:

- Around Rs 20.2 million per year are saved in the form of saving of 47,820 man-hours through this process.
- There were a total of 3985 proposals during the first year of the implementation (2017–18). In the manual process, each proposal used to consume at least 10–12 man-hours at different levels for obtaining credit letters.
- In the online system, no man-hours are lost.
- In addition to the cost of man-hours in terms of money, there is a huge advantage of the availability of people for the work at the shop floor which in turn helped in improvement in productivity. Improvement in productivity results in improvement in earnings by the organisation.
- The automated online process indirectly helped in entry and exit gate control. In the manual process, free movement was allowed in the name of the sensitivity of health concern.

7 Conclusion

Albert Einstein says ‘Everything should be made as simple as possible but not simpler’, Stephen Hawking believes that ‘Intelligence is the ability to adapt to change.’ and Jack Welch says “No company, small or large, can win over the long run without energized employees who believe in the mission and understand how to achieve it.”

The combination of all the above three statements indicates that the organisations should ensure that its employees could use the automated processes, by eliminating time delays and the simplification of procedures made possible by ICT. The success of RINL’s medical referral process in improving the morale of the employees paved the way for imbibing ICT in other processes of HR management also.

References

1. Gomes, S.: Labour welfare and industrial health facilities of labour welfare, a study on impact of non statutory welfare facilities on employees performance at reliance KG D6, Kakinada (A.P). Int. J. Res. Cult. Soc. **1**(08) (2017). ISSN: 2456-6683
2. ICT: An Effective Tool in Human Development Kuyoro Shade O. Awodele O. Okolie Samuel O. Department of Computer Science Babcock University, Ilisan-Remo Nigeria. Int. J. Humanit. Soc. Sci. **2**(7) (2012)
3. Impact of ICT on Human Resource Development, October 4 2012. <http://egov.eletsonline.com/2012/10/impact-of-ict-on-human-resource-development/>
4. RINL application for IIM award 2016-17 Published by CSM department of RINL (2017)
5. ICT and HRM: Management perceptions about... (PDF Download Available). Available at https://www.researchgate.net/publication/297918571_ICT_and_HRM_Management_perceptions_about_the_changing_scenario. Accessed 09 May 2018
6. Impact of Information and Communication Technology in HRM, 1Abhishek Vohra, 2Ankit Srivastava, 3Rohit Premi, 4Dr. Shine David MBA HR (SEM 2), Institute Of Management Studies. (DAVV), India. Int. J. Comput. Sci. Inf. Technol. Res. **3**(2), 511–516 Apr–June 2015. Available at: www.researchpublish.com. ISSN 2348-120X

Smart City and Challenges



Dinesh Kumar Saini and B. Y. Sandhiyaa

Abstract The rapid urbanization and ever-increasing population in our country give rise to impediments to the true spirit of civilization. Improper urban and rural infrastructures, growing crime rates, inadequate electricity, confusing and ill-adapted tax regulations, speculative banking, unstable policies, industrial sickness, women empowerment, restrictive labor regulations, sanitation, and non-transparent governance are some of the many issues which show that the unprecedented technological development has not kept pace with the population explosion, especially in cities. This lack of coordination is due to the scenario that the simple introduction of technological services is not sufficient to make them be used efficiently by the citizens. Therefore, there is a call for a better organization and implementation of tools of urbanization provided to the inhabitants of the country. Smarter solutions need to be executed to make digital services for economic and social advancement seamlessly reach the occupants in an easy and secure way and encourage them to continue using the amenities. A holistic development rather than just technological advancement is essential for the betterment of urban lifestyle of the present and future population. This paper attempts to analyze one such advancement, i.e., smart cities and the challenges involved in implementing them in various sectors.

Keywords Smart cities · Internet of things · Blockchain

1 The New Form of Advancement

A smart city is where services in sectors like government, health, energy, transport, banking, commerce, corporate, and education are digitized while connecting and

D. K. Saini
Sohar University, Sohar, Oman
e-mail: dinesh@soharuni.edu.om

B. Y. Sandhiyaa (✉)
SRM Institute of Science & Technology, Chennai, Tamil Nadu, India
e-mail: sandhiyabalu@gmail.com

centrally managing elements in each sector using information and communication technology (ICT). Smart cities involve the complex integration of technology, information, and a compendious application of information resources. This integration may be within and/or between them [1].

Smart city, as defined by the European Parliament (2014), is a place where the traditional networks and services are made more efficient with the use of digital and telecommunication technologies, for the benefits of its inhabitants and businesses [2].

Smart city technology rigidly connects the administration to the community and offers the citizens a more approachable and accountable system of governance [3]. It helps enhance the quality of the city infrastructure, reduce usage of environmental capital, make resource consumption sustainable, and improve city operations and service deliveries.

2 Smart City Design

Designing smart cities for today's urban infrastructure requires a collection of various kinds of data ranging from governmental policies and regulations to common people's needs. For a user-friendly digital public service delivery, smart city plans should be strengthened with improved enthusiasm of the officials and capabilities of the administration, policies that assure personal data encryption, and insurance against severe service failures, reference successful models, architecture, and design ideas to efficiently deliver co-created services.

2.1 *Challenges Before Initiating Plan*

There are challenges that persist as a hurdle to the initiation process of designing a smart city. The cost of reconstructing a city's infrastructure would be higher than building a new one. The lack of finances for supporting this transformation is very a critical issue, especially in developing countries. In huge democracies like India, this idea of a big change might not be welcomed by a large part of the population due to various reasons like the possibility of life becoming expensive in the cities and diminution of their heritage. Without citizen participation, the design and policy-making process would be incomplete and meaningless [4].

These challenges can be addressed by making a solid foolproof plan of architecture and implementation and keeping the population informed about the development of the smart city and its benefits and involved in the reconstruction. This could be done through various online platforms, social media, public forums, etc. It is important to keep the citizens aware throughout the process that it is a nonprofit initiative taken solely for increasing the quality of life for everyone in the area. (The people must be constantly motivated to be eager to utilize the smart system for their best

benefit.) It is also essential that the hierarchy of officials involved is clear, sturdy, and confusion-free.

2.2 Preparation for Plan

The planning should be begun by pooling in data on various needs of the population and various services that remain inefficient or unutilized. Then, generating various technologically implementable solutions to existing and potential problems. This can be done using crowdsourcing. Crowdsourcing is a technique by which a group of people collaborates through Internet-based platforms to gather ideas, support, content, and innovative solutions [5]. It is an excellent method to elevate public engagement.

The ideas, data, and solutions thus exhaustively collected may be further analyzed, inferred, and organized. Since the data collected are from a group of various sectors of the community, it should be translated such that it is of relevance to the intended context; i.e., the raw data should be classified into different aspects of smart city planning in order to be used as information. For better understanding and reducing ambiguity, the analysis process could be aided by diagrammatic representation of various ideas and categories using flowcharts, graphs, and mind maps [6]. This qualitative content analysis is unfinished without various checks for validity, accuracy, reliability, rationality.

The outcome of this phase is generally a well-ordered set of services, policies, and efficient digital solutions to execute them.

2.3 Resource Optimization

One of the major objectives of the smart city technology is the transformation from automation to optimization. Optimization is the process of achieving desired results or maximizing/minimizing the output with minimum input. In today's world of population proliferation, resource optimization is the smarter solution to a sustainable lifestyle. A smart city resource could be many things: capital, energy, transport, manpower, land, technology, time, and many more.

Optimization problems are generally represented as

Minimize

$$f(x) : R^n \rightarrow R \quad (1)$$

subject to

$$h_i(x) = 0, (i = 1, 2 \dots .I) \quad (2)$$

$$g_j(x) = 0, (j = 1, 2 \dots J) \quad (3)$$

where h_i and g_j are called equality and inequality constraints, respectively [7]. This representation is defined for a minimization problem. By negating $f(x)$, it can be used for a maximization problems.

Optimal solutions are conventionally calculated using linear programming, quadratic programming, algebra, convex optimization, and other mathematical algorithms.

Optimization problems are mostly nonlinear, multimodal, and extremely complex. Hence, traditional algorithms are less preferred for optimization. Traditional algorithms do not always guarantee global optimality.

Resource optimization problems in smart cities deal with massive amounts of computational data and constantly varying environment of implementation. Therefore, more heuristic methods of approach are vital.

Metaheuristic algorithms, advanced heuristics that use memory, solution history, and other complex learning processes are most suitable for RO problems as they are not problem-specific, most likely to find optimal solutions. Their stochastic (statistically random) property is very beneficial as no pattern in a smart city technology can be precisely predicted.

Since our problem predominantly involves the simulation of various human activities and interactions, naturally inspired metaheuristic algorithms best serve the purpose [8].

These algorithms help cope with the increasing dimensions, variables, time, and space complexity, of engineering problems.

Nature-inspired algorithms involve large populations of different agents like bats, ants, bees, and fireflies. Different solutions are evolved with different properties (Fig. 1).

2.4 Measurement

Anything that is to be developed needs to be measured. Smart city technology's implementation involves the improvement of different aspects of an area like infrastructure, service availability, land, networking facilities, and devices/machines to be connected. The challenge is that it needs to be appropriate to the city ratios. Measurements are therefore important to design, deploy, and track this advancement. Optimization of resources can be carried out effectively only if there is accurate data about the current urban scenario. Several qualities of life factors that need to be measured include population, finance, economy, social engagement, security, urban planning, geography, climate, technology availability, and transportation [9].

Although it is a process that potentially brings out the best output possible, there are challenges in measuring all aspects of a system as complex as smoothly operating city.

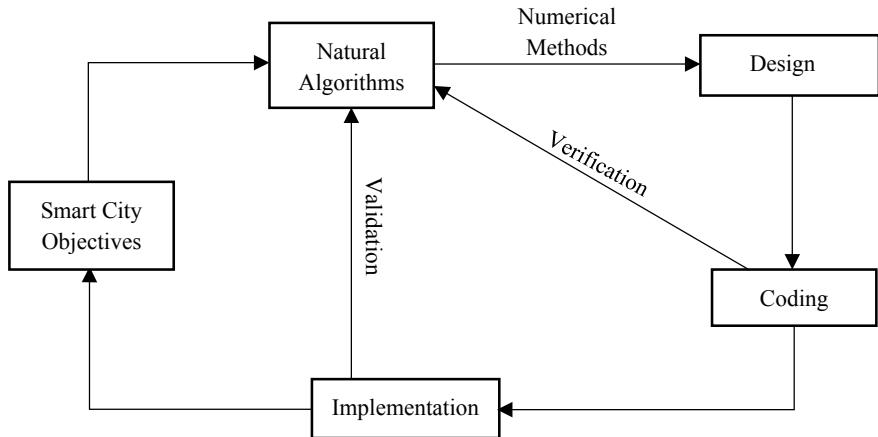


Fig. 1 Resource optimization using natural algorithms

One major difficulty is the selection of correct set of metrics and measurement tools for the set of factors because of the uniqueness of each factor. The city under consideration should be studied carefully before finalizing the mensuration and data collection methods.

2.5 Implementation

The concept of making cities smarter is to focus on the drawbacks of mere technological advancement. Digitizing services has not been quite in pace with their proper utilization and availability, therefore not meeting to their true purposes.

Internet of Things for Availability of Technology for Use. The key components of this new method of urban development include more than just smart technologies. This concept proposes to make everything around us smarter. When the things around us like traffic signals, cars, mobile phones, and solar panels are connected to each other and centrally monitored and managed, the processed information and service delivery are readily available for utilization. Improved and intelligent algorithms and systems are essential to visualize, process, and deploy the vast amount of information generated from these smart things. This is where the Internet of things technology comes to the rescue.

These smart devices communicate through the network and enable data transfer. The data collected from each device are stored and delivered back in the form of processed information or a service not only to that device, but any device connected in the network. All the raw data are stored, managed, and processed in the Internet using cloud computing technology [10].

Blockchain for Security. When almost every element of a city becomes a part of an ever-working network of data transfer and processing, a lot of information is made public. Such transparency questions the security of the interconnected system. Smart cities generate security threats that are unmatched by any other technological advancement. A system interconnecting a large number of gadgets and people poses dangers ranging from social surveillance to physical security threats. Safety cannot be considered an afterthought when it comes to redesigning an entire city. Despite the availability of several technical solutions for cybersecurity issues, the vulnerability of interconnection of so many elements of such a large area cannot be addressed [11].

Since smart city security is a bigger challenge than ordinary tech-security, we need smarter solution. The newly developed concept of block chains may be utilized for this purpose.

Initially described by Stuart Haber and W. Scott Stornetta in 1991 and actually conceptualized by Satoshi Nakamoto (developer of bitcoin) in 2008, blockchain is the concept of cryptographically secured chain of blocks. It is a public record of transactions which is distributed among various computers. After each valid transaction, the record is updated by linking it to the previous transaction.

The security of this record is ensured because to modify them one has to break into thousands of computers involved in validating the transactions.

The validation is based on smart contracts which are simple pieces of code that are added to the blockchain and hence cannot be manipulated. Smart contracts can be written by anyone. However, once written the programmed behavior of the contract, it cannot be changed. The only way to change it is by adding a new smart contract to the chain provided everyone agrees to it. The previously written one remains ceased to be used yet remains in the blockchain unmodified. This makes blockchain an incorruptible and reliable means of conducting all kinds of transactions.

There is the added benefit of reduced cost because of the shared public ledger which makes auditing records easier.

Enabled by the blockchain concept, trust-free cryptographic data transactions can be done among the IoT devices within and among the smart cities [12].

3 Future Scope of the Research

This paper was a general investigation on the challenges to smart cities and an attempt to propose upcoming technologies that could be used to solve those shortcomings. This research would be continued, and more specific methods of implementation would be analyzed and documented in the future extension of this paper. Technical data collection for more comparative and graphical analysis has been initiated. Thus, they can be expected in the more detailed extension of this research.

References

1. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart cities in Europe. *J urban Technol.* **18**(2), 65–82 (2011)
2. Manville, C., et al.: Mapping smart cities in the EU (2014)
3. Paskaleva, K.A.: Enabling the smart city: The progress of city e-governance in Europe. *Int. J. Innov. Reg. Dev.* **1**(4), 405–422 (2009)
4. Nam, T., Pardo, T.A.: Smart city as urban innovation: focusing on management, policy, and context. In: Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance. ACM (2011)
5. Schuurman, D., et al.: Smart ideas for smart cities: investigating crowdsourcing for generating and selecting ideas for ICT innovation in a city context. *J. Theor. applied electronic commerce research* **7**(3), 49–62 (2012)
6. Cardone, G., et al.: Fostering participation in smart cities: a geo-social crowdsensing platform. *IEEE Communications Mag.* **51**(6), 112–119 (2013)
7. Koziel, S., Yang, X.-S. (eds.): Computational Optimization, Methods and Algorithms, Vol. 356. Springer (2011)
8. Yang, X.-S.: Nature-Inspired Metaheuristic Algorithms, Luniver press (2010)
9. Neirotti, P., et al.: Current trends in smart city initiatives: some stylised facts. *Cities* **38**, 25–36 (2014)
10. Avijit, K., Chinnaian, R.: IOT for Smart Cities (2018)
11. Elmaghraby, A.S., Losavio, M.M.: Cyber security challenges in smart cities: safety, security and privacy. *J. Adv. Res.* **5**(4), 491–497 (2014)
12. Biswas, K., Muthukumarasamy, V.: Securing smart cities using blockchain technology. In: High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016 IEEE 18th International Conference on, IEEE (2016)

Flow Shop Scheduling in Which Processing Time Connected with Probabilities and Job Delay Due to Maintenance for $M * N$ Machine



E. Janaki and A. Mohamed Ismail

Abstract The objective of the paper is to obtain optimal or close to optimal makespan for $m * n$ sequential problem, wherein processing time connected with their weighted probabilities. In this paper, we analyse the processing time interval of the job may or may not have the maintenance activity and breakdown. We construct a job sequence by heuristic method and optimal makespan and also obtained mean flow time. The method is illustrated by using some numerical examples.

Keywords Flow shop scheduling · Job delay due to maintenance · Mean · Flow time · Makespan and heuristic method

1 Introduction

A flow shop scheduling problem exists when every job shares a similar handling request on every machine. The flow shop contains m machines and set of n jobs are to be processed. In flow shop problem, we take that each job has different processes. A flow shop is described by same directional flow of work with the assortment of job being handled consecutively one way. The processing time related to their probabilities of considerable number of jobs is thought to be known and every job is handled in a similar request in various machines. Researches from different fields are pulled into the investigation of scheduling problem. The sequencing problem is considered with deciding the sequence in which a set of jobs is to be processed by everyone of the machine. Good scheduling is necessary to maintain system effectiveness and control of task. In scheduling, there are many scenarios for handling unavailability intervals. Any scheduling problem basically relies on three vital factors transportation time, relative significance of the job over another job and breakdown machine. These three elements were advised by numerous researchers. Miyazaki and Nishiyama (1980) had investigation for minimizing weighed mean flow time in flow shop scheduling

E. Janaki (✉) · A. Mohamed Ismail
Sathyabama Institute of Science and Technology, Chennai, India
e-mail: januelango@gmail.com

[1]. Chandramouli (2005) proposed a heuristic approach for n job, three-machine flow shop scheduling problem including transportation time, breakdown time and weights of jobs. Pandian and Rajendran (2010) enhanced and simplified the methodology for a constrained FSP for three jobs [2, 3]. Optimal three-phase production schedule the processing time and set-up times associated with probabilities including job block criteria is discussed by Singh [4]. Having analysed different situations, in this paper, we propose a new model for flow shop scheduling problems involving known breakdown time, weights of jobs and three different types of processes with n jobs and m centre. The initial optimal solution is to be obtained by considering the weights using any simple heuristic method. Then, without modifying the sequence, the makespan and weighed average time can be optimized after imposing the breakdown time constraints. The n jobs m machines problem, the method proposed by Johnson (1954) can be used effectively [5]. Other heuristics include the slope index (SI) method proposed by palmer (1965), functional algorithm proposed by Gupta (1975) [6, 7]. Both CDS and RA (rapid access procedure) heuristics proposed by Campbell et al. (1970) and Dannenbring (1977), respectively, are used on Johnson's algorithm for the two-machine problem and are simple to use [8, 9]. Six decades of flow shop scheduling were analysed by Tyagi (2013) [10]. A Simple model to optimize general FlowShop scheduling problems with known break down time and weights of jobs is discussed by Baskar and Anthony Xavior [11]. A heuristic algorithm proposed method is very simple and easy to understand and also, can be used as a tool for flow shop scheduling problems. With the help of a numerical example, the proposed method is illustrated.

2 Assumptions

1. Initially, all the jobs are ready to go in the machine.
2. Set-up times for the machines are included in processing time.
3. Second-machine job is permitted only after completion of job on the first machine.
4. Idle time can be considered.
5. Constraint factors are known in advance.
6. Job can Delay due to certain temperature.
7. Consider n jobs say $i = 1, 2, 3 \dots n$ are processed on four machines M_1, M_2, M_3 and M_4 and processing times may be classified under three cases:
 - (i) Machine-1 (M_1)—need not consider maintenance time since it is a new machine. Maintenance time will be ignored.
 - (ii) Machine 2 & 3 (M_2, M_3)—the process in these machines can be continued if any maintenance activity is performed in between the particular process. Maintenance time will be considered.
 - (iii) Machine 4 (M_4)—the process in this machine cannot be continued if any maintenance activity is performed in between the particular process. Maintenance time will be considered.

8. The job to be assigned with a weight w_i , mean flow time is defined by:

$$\frac{\sum f_i w_i}{\sum w_i}$$

3 Mathematical Illustration

Consider n jobs say $i = 1, 2, 3 \dots n$ are processed on four machines M_1, M_2, M_3 and M_4 with the processing time a_i, b_i, c_i and d_i , respectively, and their corresponding probabilities p_i, q_i, r_i and u_i . Table 1 gives the list of machine and their processing time in the notational values.

3.1 Algorithm

Step 1: Calculate expected processing time on machines on four machines

- (i) $A_i = a_i \times p_i$
- (ii) $B_i = b_i \times q_i$
- (iii) $C_i = c_i \times r_i$
- (iv) $D_i = d_i \times u_i$

Step 2: Using Palmer's approach, optimal sequential order is derived

Step 3: Identify the breakdown interval by the known value

Step 4: Now categorize the jobs and modify the processing times suitably by the following cases,

t —Existing processing time

t_{new} —New processing time

a, b —Break down time span start and end

S_1, S_2 —Existing processing time span start and end

Table 1 Multiple machines and their processing time

Job	M_1		M_2		M_3		M_4		Weight
I	a_i	p_i	b_i	q_i	c_i	r_i	d_i	u_i	w_i
1	a_1	p_1	b_1	q_1	c_1	r_1	d_1	u_1	w_1
2	a_2	p_2	b_2	q_2	c_2	r_2	d_2	u_2	w_2
—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—
N	a_n	p_n	b_n	q_n	c_n	r_n	d_n	u_n	w_n

Table 2 Sample data

Job	M_1		M_2		M_3		M_4		Weight
I	a_i	p_i	b_i	q_i	c_i	r_i	d_i	t_i	w_i
1	70	0.3	20	0.1	50	0.2	80	0.2	4
2	60	0.2	30	0.2	90	0.3	40	0.1	3
3	40	0.2	10	0.1	50	0.2	70	0.2	2
4	20	0.1	40	0.3	60	0.2	50	0.2	2
5	60	0.2	40	0.3	40	0.1	90	0.3	5

Table 3 Expected processing time for machines M_1 , M_2 , M_3 & M_4

Job	M_1	M_2	M_3	M_4	Weight
I	A_i	B_i	C_i	D_i	w_i
1	21	2	10	16	4
2	12	6	27	4	3
3	8	1	10	14	2
4	2	12	12	10	2
5	12	12	4	27	5

Case (i): If the process is a continuous without interruption then find new existing processing time add $(b - S_1)$ to the existing processing time,

Case (ii): If process is not continuous unaffected by any interruption, then find new existing processing time with the breakdown criteria,

Starts in between, starts and end in between, ends in between the processing time then add $(b - a)$, $(b - a)$ and $(b - S_1)$ to the existing processing time, respectively.

Case (iii): Finding new processing time for process which does not require maintenance activity should be ignored.

Step 5: If the job is affected by the breakdown, then the process to be iterated.

3.2 Numerical Example

Step 1: Consider five jobs on four machine flow shop problem whose processing time of the jobs associated with their probabilities. The following Table 2 gives a sample data of five jobs on four machines.

Solution: As per step 1, the expected processing times for machines M_1 , M_2 , M_3 and M_4 are provided in the following Table 3.

Step 2: By heuristic algorithm, namely palmer's approach,

$$A_j = \sum_{i=1}^m (m - (2i - 1)) p_{ij}$$

where p_{ij} represents the processing time of i th machine of j th job

Table 4 Optimal sequence of given job

Job	M_1	M_2	M_3	M_4	A_j	Weight
<i>I</i>	A_i	B_i	C_i	D_i		w_i
1	21	2	10	16	-7	4
2	12	6	27	4	-3	3
3	8	1	10	14	27	2
4	2	12	12	10	24	2
5	12	12	4	27	37	5

Table 5 Machine processing time interval

Job	M_1	M_2	M_3	M_4	Weight
<i>I</i>	A_i	B_i	C_i	D_i	w_i
5	0–12	12–24	24–28	28–55	5
3	12–20	24–25	28–38	55–69	2
4	20–22	25–37	38–50	69–79	2
2	22–34	37–43	50–77	79–83	3
1	34–55	55–57	77–87	87–103	4

Hence, the optimal sequence can be computed as 5–3–4–2–1. The makespan is computed as shown in Table 4.

Optimal makespan of the given jobs on the machine = 103 units.

Mean weighted flow time = 60.4 units.

The breakdown interval being $(a, b) = (29–35)$.

From the above Table 5, all the jobs are affected by the breakdown intervals.

Affected Jobs are:

Job 5 : $M_4(28–55)$

Job 3 : $M_3(28–38)$

Job 4 : $M_2(25–37)$

Job 2 : $M_1(22–34)$

Job 1 : $M_1(34–55)$

Using the guidelines, the identified processing times that is to be modified based on cases shown in Table 6:

The modified processing time is stated below in Table 7.

In-out table for the optimal sequence considering breakdown time (Table 8).

Makespan of the given jobs on their respective machine = 110 units.

The breakdown interval being $(a, b) = (29–35)$

The processing time spans for no other jobs are affected by breakdown, and furthermore iterations are not required.

The updated optimal sequence with jobs are shown in the below Gantt chart (Fig. 1).

Table 6 New existing processing time

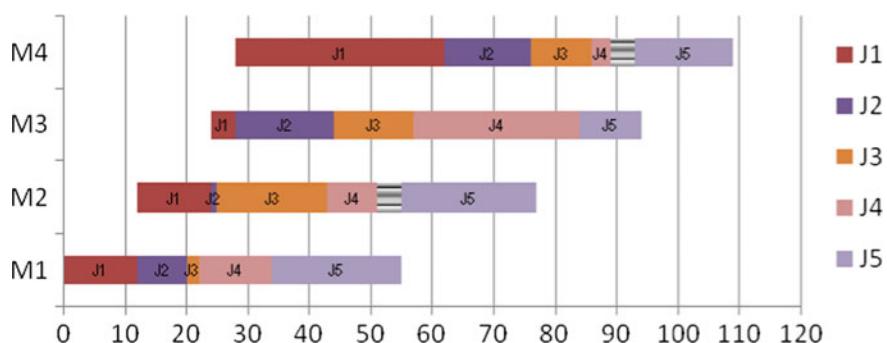
Job	Machine	Breakdown	New existing processing time		
			Start	End	t_{new}
1	M_1	X	✓		$t_{\text{new}} = t$
2	M_1	✓		X	$t_{\text{new}} = t$
3	M_3	✓	✓		$t + (b-a) = 16$
4	M_2	✓	✓		$t + (b-a) = 18$
5	M_4	✓	✓		$t + (b-s_1) = 34$

Table 7 Modified processing time

Job	M_1	M_2	M_3	M_4	Weight
I	A_i	B_i	C_i	D_i	w_i
5	12	12	4	34	4
3	8	1	16	14	3
4	2	18	13	10	2
2	12	8	27	4	2
1	21	22	10	16	5

Table 8 In-out table

Job	M_1	M_2	M_3	M_4	Weight
I	A_i	B_i	C_i	D_i	w_i
5	0–12	12–24	24–28	28–62	4
3	12–20	24–25	28–44	62–76	3
4	20–22	25–43	44–57	76–86	2
2	22–34	43–51	57–84	86–90	2
1	34–55	55–77	84–94	94–110	5

**Fig. 1** Gantt chart

4 Conclusion

The objective was to find the optimal makespan for $m * n$ machines and also considering the breakdown in supply criteria. This work may be extended with different constraints such as set-up time and transportation time with fuzzy environment. This can also be extended to stochastic processing time.

References

1. Miyazaki, S., Nishiyama, N.: Analysis for minimizing weighted mean flow-shop scheduling. *JOR Soc Jpn* **23**, 118–132 (2000)
2. Chandramouli, A.B.: Heuristic approach for n-job, 3-machine flow-shop scheduling problem involving transportation time, breakdown time and weighs of jobs. *Math. Comput. Appl.* **10**, 301–305 (2005)
3. Pandian, P., Rajendran, P.: Solving constrained flow shop scheduling problems with 3 machines. *Int. J. contemp. Math. Sci.* **10**(19), 921–929 (2010)
4. Singh, T.P., Rajendra, K., Gupta D.: Optimal three stage production schedule the processing time and set up times associated with probabilities including job block criteria. In: Proceeding of National Conference FACM, pp. 463–470 (2005)
5. Johnson, S.M.: Optimal two- and three-stage production schedules with setup times included. *Naval Res. Log.* **1**(1), 61–68 (1954)
6. Palmer, D.S.: Sequencing jobs through a multi-stage process in the minimal total time-a quick method of obtaining a near optimum. *Oper. Res.* **16**, 101–107 (1965)
7. Gupta, J.N.D.: A functional heuristic algorithm for the flow shop scheduling problem. *Oper. Res.* **22**, 39–47 (1971)
8. Campbell, H.G., Dudek, R.A., Smith, M.L.: A heuristic algorithm for n job m machine sequencing problem. *Manage. Sci.* **16**, 630–637 (1970)
9. Dannenbring, D.G.: An evaluation of flowshop sequencing heuristics. *Manage. Sci.* **23**, 1174–1182 (1977)
10. Tyagi, N.: Six decades of flowshop scheduling research. *J. Sci. Eng. Res.* **4**, 854–864 (2013)
11. Baskar, A., Anthony Xavior, M.: A simple model to optimize general flow—shop scheduling problems with known break down time and weights of jobs. *Precedia Eng.* **38**, 191–196 (2012)

IoE-Enabled Smart Embedded System: An Innovative Way of Learning



Akshada Rathod, Prachi Ayare, Ramchandra Bobhate,
Rajneeshkaur Sachdeo, Sambhaji Sarode and Jyoti Malhotra

Abstract In modern days, there is a need for an effective, smart, and quick teaching methodologies helping students to have intelligent and thorough knowledge experience. Effective teaching depends on deliverable methods and moral learning environment. New trends in this regard have raised the term Internet of everything (IoE) which connects machine, people, and processes. This paper offers an IoE-based solution to upgrade teaching and learning processes. IoE covers the entire digital age with miniaturized operating system devices serving easy integration of large applications on these tiny devices. We present an innovative approach for creating a smart learning environment, in which teaching resources are automatically operated after entering credentials through a Web application. Learning strategy also supports recording functionality wherein teacher can record the lecture contents and students can view recorded lecture series and can also interact through the collaborative screen. All modules are being controlled through embedded computing device Raspberry Pi for hardware interfacing, Web application, intelligent media for controlling the resources, and scripting module for the collaborative screen. As different features are integrated into a centralized zone along with a modest and easily operated interface, this system results as a worthwhile and helpful for both teachers and students.

A. Rathod (✉) · P. Ayare · R. Bobhate · R. Sachdeo · S. Sarode · J. Malhotra

Department of CSE, MIT School of Engineering, ADT University, Rajbaug,
Loni-Kalbhor, Pune 412201, India

e-mail: akshadarathod2@gmail.com

P. Ayare

e-mail: prachi.ayare@gmail.com

R. Bobhate

e-mail: gbobhate@gmail.com

R. Sachdeo

e-mail: rajni.sachdeo@mituniversity.edu.in

S. Sarode

e-mail: sambhaji.sarode@mituniversity.edu.in

J. Malhotra

e-mail: jyoti.malhotra@mituniversity.edu.in

Keywords IoE · Collaborative screen · Raspberry Pi · Interfacing · Smart learning

1 Introduction

Internet of things refers to connecting physical objects together, which are capable of sensing the context of the surrounding environment and to exchange the collected information. Change in technology has transformed the term Internet of things (IoT) to the Internet of everything (IoE). IoE is interconnecting people, processes, things, and data together so that they can communicate intelligently with each other without human intervention. The IoE-based application is generally related to real-life entity module in which every device is connected and managed by a central controller called as module controller. Key components of IoE comprise sensor devices, controlling devices, and data management along with runtime applications [1, 2]. Every IoE application is classified into two types: intranet system and extranet system. Today, it is used in various fields such as agriculture, automobile industry, power management, water management, environmental monitoring, healthcare industry, transport system, and many more to automate every object by consuming less energy and more revenue. VSRS [3] communication protocol can be implied for low-power IoT-enabled sensing devices to achieve reliability based on their priority index. IoT has impacted people's lives by providing personal safety, saving time, and better utilization of natural resources.

One of the most important applications is the education sector. The quality of learning and teaching is based on content and deliverable methods. These deliverables are also changing with the evolution of IoT leading to the smart learning environment.

The traditional learning environment has some open issues, wherein manual help is required in operating resources such as a projector and a public address (PA) system. IoE can provide an automatic operating mechanism for these resources along with other teaching facilities like flip the classroom, collaborative screen sharing, and recording the lectures. The collaborative screen can be used in group discussion and problem-solving where students can participate in being at a remote location.

This paper presents innovate way of learning, i.e., IoE-enabled smart embedded system (IoE-SES) in automating a few teaching resources. The key contributions comprise three major modules: 1. *Automate* the learning environment by controlling the projector and sound system remotely using a Web server; 2. *Create* a real-time collaborative screen for the multiuser environment; 3. *Record* lecture videos.

Further sections cover the related work, working methodology, output, followed by a conclusion and future scope.

2 Literature Review

Vaidya et al. [4] presented a smart automation solution for oldies, wherein home appliances are managed by an embedded programmed device operating on mobile user commands. Concerning security and energy efficiency, appliances are controlled using face and voice detection. Detection accuracy is measured in terms of acceptance and rejection ratio. Menezes et al. [5] proposed a system to detect and monitor entities in a surveillance area using an open-source framework and Raspberry Pi by switching on the lights and taking surrounding images. As the captured image is authenticated and encrypted at both ends, then the system can be used in restricted areas that reduce the cost required for surveillance. Lamine et al. [6] developed a mechanism to control multiple home appliances using an android application and a Web server. The system is implemented using a representational state transfer (REST) Web services to handle different devices remotely and efficiently adjusting the height of the garage shutter. Patchava et al. [7] demonstrated an affordable surveillance system for smart home automation using Python programming language. Users can interact with the system using mobile devices and control electronic appliances to get the current status of devices. Video live streaming is done through the Internet for sensing and surveillance. In this system, users can check and update the current status of devices connected by Raspberry Pi. Over usage and power notifications will be sent to the users. Consequently, the further necessary decision can be taken according to power consumption situations. Sandeep et al. [8] designed a home appliance remote controlling system with higher mobility and security using weaved cloud services. The proposed system is built using a Raspberry Pi embedded devices with an electromagnetic relay for power regulation. Authors compared the features of the proposed system with traditional systems. However, it differs in terms of the transmission of unique code through Web service. Thus, it reduces the network overheads to control each device separately. Zhang et al. [9] proposed a socket program using Node.js to build a Web server based on an express framework with MongoDB. It is used to establish communication between Web applications and embedded systems. This approach uses the model view controller (MVC) architecture with hypertext transfer protocol (HTTP) request and response model. Security is achieved using a cyclic redundancy check (CRC). Stergiou et al. [10] surveyed different papers of cloud computing and Internet of things by taking the security aspects into consideration. The authors have also integrated both technologies together to determine advantages and their common features. In addition, RSA and advanced encryption standard (AES) security algorithms for cloud computing are discussed with networking parameters. Xue et al. [11] proposed a coding solution for videoconferencing system with improved screen sharing using Web real-time communication (WebRTC) technology. Node.js was used to set up connection and to transmit alert messages. High-resolution videoconferencing is achieved even in a low-bandwidth network. The comparative study of the work discussed here with the proposed work is summarized in Table 1.

Table 1 Analysis and comparative study of IoT system parameters

Parameters	SHA-UDMS [4]	SMS [5]	SHAT [7]	W-VCSSS [11]	IoE-SES (Proposed)
Development board	RPi, Arduino	RPi	RPi 3	RPi 3	RPi 3B
Development framework	OpenCV	SimpleCV, MPJG streamer	NA	MVC	OpenCV, MVC
Functionalities	Face capturing and recognition	Motion detection	Check the status, and control the home appliances	Videoconference. Improved screen sharing quality	Automate PA, projector. Video recording, and collective screen
Objectives	Home automation	Surveillance of the environment	Energy saving	Videoconferencing with screen sharing	Smoothen teaching and learning processes

3 IoE-SES Methodology

This segment presents the particulars of IoE-SES working procedure and system implementation as expressed in Fig. 1. The entire system is built around the multi-purpose embedded device like Raspberry Pi and a Web application to not only control different electronic devices like a projector and sound system, but also presents the collaborative screening and video recording. The teacher has an access to the user interface, where he/she can log in using radio-frequency identification (RFID) or entering a password. System access is granted for authorized users. The teacher can commence the lecture by switching IoE-enabled projector and sound systems, followed by usage of collaborative screening and recording the lecture videos. Recorded videos are initially deposited on Raspberry Pi and are further relocated to a MySQL database or to secondary storage. If required, the teacher can edit the videos as per his/her convenience. A collaborative screening can be used in group discussions, where students can mark their participation through the classroom or being at remote locations. Teachers, as well as students, can draw or write anything on the screen, which can be observed by other participants in real time.

Algorithm 1 illustrates the IoE-SES implementation initiating via a Web application. Primarily, the teacher is requested to register his/her credentials to the Web application. Post successful authorization, the teacher can gain admittance to the entire system.

Major components/cases listed in the Algorithm 1 are briefly described below:

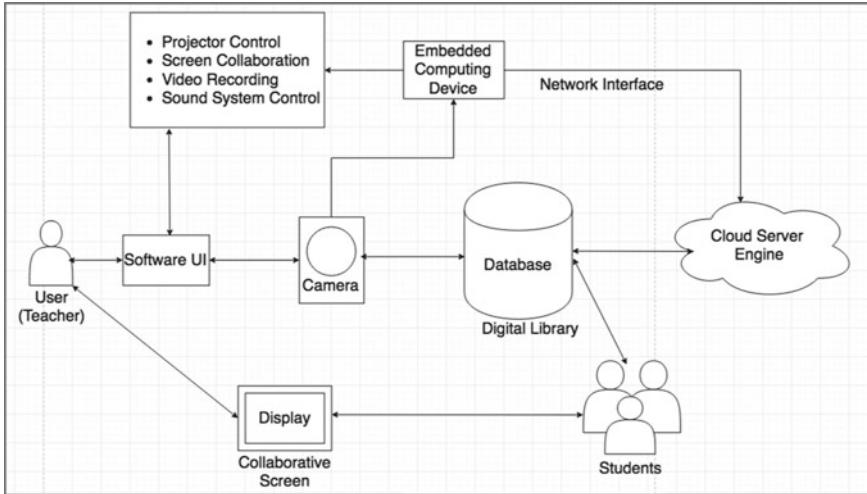


Fig. 1 IoE-SES working procedure

1. The automatic device control system

This module uses emerging IoE technology to automate the projector and sound system in the room. All these devices can be managed on a single click. Sometimes students/teachers have a tendency to forget to switch the equipment on or off. Thus, a solution is implemented to control room environment using the Web app, wherein an individual can access the projector and sound system from anywhere and anytime. This saves a lot of power consumption and thus helps in saving energy. Each device has only two states either on or off; this is defined in Eq. (1). When the user gives command on/off through user interface, GPIO pin of Raspberry Pi is either enabled or disabled resulting in switching the devices on/off. The setup is simulated in Fig. 2a, and the output is illustrated in Fig. 2b, c.

$$\emptyset(p) = \begin{cases} 1, & \text{Switch on the electronic device} \\ 0, & \text{Switch off the electronic device} \end{cases} \quad (1)$$

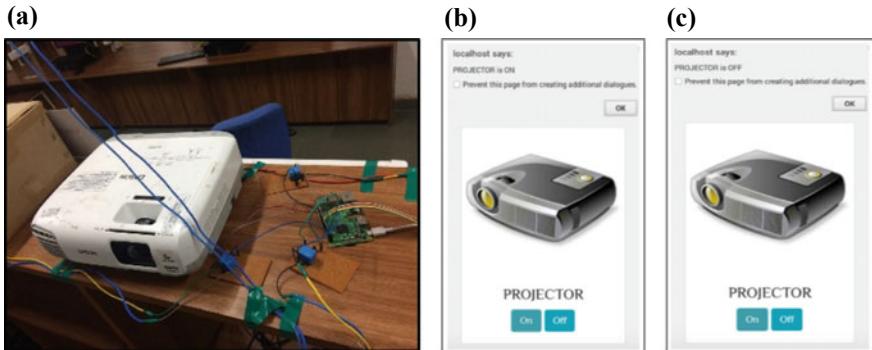


Fig. 2 **a.** Implementation of an automatic device control system, **b.** Projector ON, and **c.** Projector OFF

Algorithm 1: IoE-SES system implementation

Input: User input (Login details, operative commands)

Output: Actions (Projector, sound system – ON/OFF, video recording, collaborative screening -Start/Stop)

Begin: IoE-SES

 Connect Raspberry Pi to weaved web services to run a web application.

Case 1: Switch on/off the projector and sound system //Automatic device control

If General Purpose Input Output (GPIO) input is high **then**

 Switch the projector and sound system ON

Else if GPIO input is low **then**

 Switch the projector and sound system OFF

End if

Case 2: Share screen among the students //collaborative screening

Case 3: Video recording and uploading

 1. If command == start video capture **then**

 Capture video in .avi format and upload it

 2. Edit video to remove noise and unwanted content

 3. Upload and share the videos with students.

End

2. Collaborative screen

In a collaborative screening, the screen is shared with multiple users. Every user who has access to the screen can draw or write whatever he wants, and other users can view those changes in real time. This is achieved using Node.js (socket.io) and Web sockets. Screening module is implemented at server and client sides as captured in Fig. 3a, b.

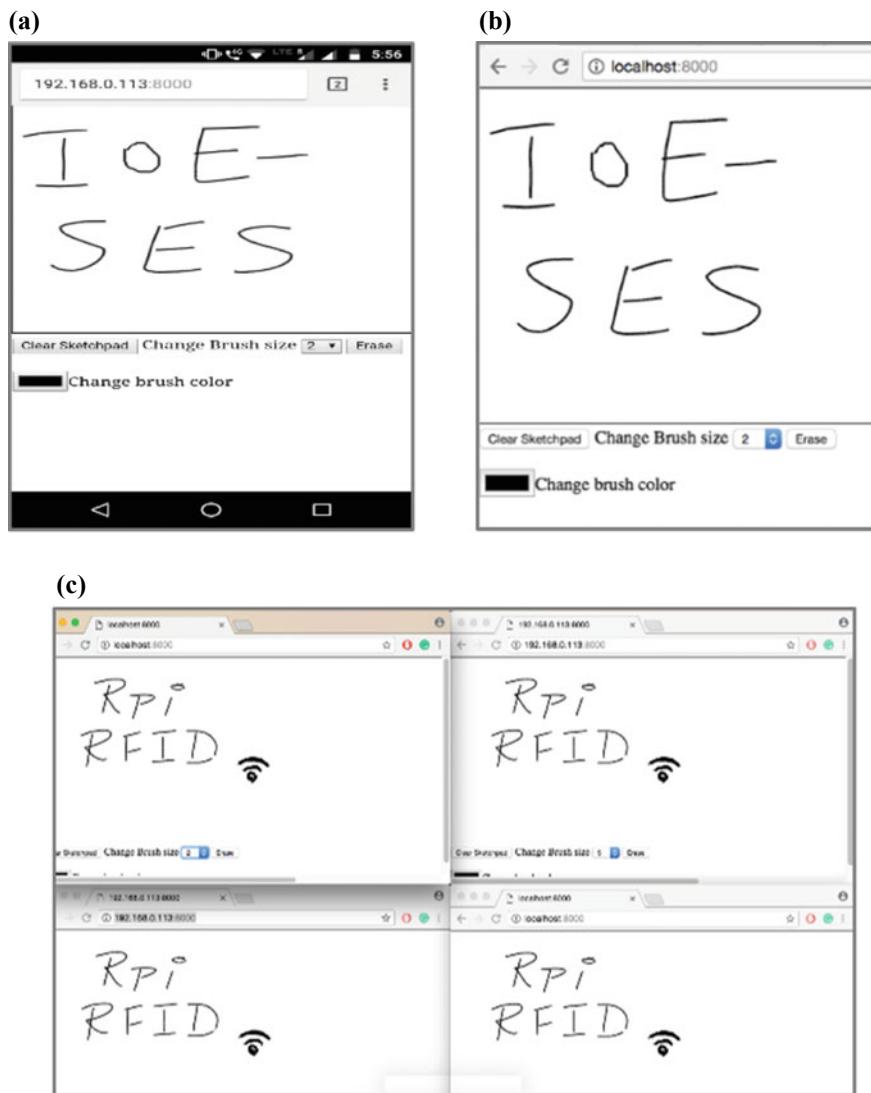


Fig. 3 **a.** Collaborative screen—server side, **b.** collaborative screen—client side, and **c.** collaborative screen—multiple browsers

Fig. 4 Implementation of video recording module



Server-side implementation	Client-side implementation
<ul style="list-style-type: none"> Initially, the Web socket server is created using Node.js and socket.io Create socket.io to listen at port 8000 Compile and run the node and initiate server on port 8000 as shown in Fig. 3a 	<ul style="list-style-type: none"> Create a canvas element for drawing or writing Add draw function to the canvas and drag the required events Create Web socket at port 8000 Post connection establishment, the system draw events as shown in Fig. 3b Data is sent to all open browsers as presented in Fig. 3c

3. Video recording and uploading

The effective teaching–learning process is achieved using video recording and sharing the content over the Web. Shared videos can be referred by students later for revising the concepts. In order to achieve this, the system contributes to video recording, content editing, and sharing. Duplicate videos are eliminated using the deduplication method projected in [12]. In addition to this, students can ask queries to the concern teachers over Question Answer (QA) forum. This not only saves the time of offline teaching but also helps to reduce the gap of communication between students and teacher. Implementation of this module is simulated in Fig. 4.

Furthermore, to achieve this, a secure authorization mechanism is implemented using either RFID NFC technology or login facility. When lecture begins, the teacher can start recording the lecture by capturing the video through the graphical user interface. Moreover, recorded videos are processed using OpenCV. They are stored in MySQL database for editing purpose. Later they are shared with students over the Internet. Recorded videos are always depended on frames that are captured. Equation (2) shows that the time required to capture a video depends on a number of frames captured per second.

$$\forall(\text{Recorded Videos}) \propto (\text{Number of frames}) \quad (2)$$

Total time required to upload a video in the database is given by Eq. (3).

$$T = \partial(t_r) + \partial(t_u^s) * F \quad (3)$$

where T = total time; F = total number of frames

$\partial(t_r)$ = time require to record video

$\partial(t_u^s)$ = time require to upload one frame of video to the server

Algorithm 2: Video capturing

Input: User input (Video start/stop)

Output: Video in .avi format //audio video interleave format

Begin: video_capture

 Import required libraries and Initialize object of video_capture()

While recording the video

 capture frames, save in .avi format; and upload it

End while

 Release all allocated resources;

End

4. Web application development

This Web application acts as an intermediary between Raspberry Pi and electronic devices. It is developed in PHP technology, designed in Bootstrap style, and hosted by weaved cloud services. This application gives access to the system by unlocking it using RFID reader or by entering username and password.

4 Conclusion and Future Scope

We have successfully implemented the automation of projector and sound system using Raspberry Pi. The Web application can be accessed by login using RFID as well as entering a password. Videos can be recorded and stored on secondary storage. Currently, the system is using RFID for authorization, but more security aspects like authentication can be added to make the system more secure. Recorded videos can be stored and edited on clouds instead of storing them on Raspberry Pi so that authorized persons can access them, and Raspberry Pi can be used more efficiently.

References

1. Hussain, F.: Internet of things: “Building blocks and business models”. No. 978-3. Springer International Publishing (2017)
2. Lee, I., Lee, K.: The internet of things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz. **58**(4), 431–440 (2015)
3. Sarode, S., Bakal, J.: VSRS: variable service rate scheduler for low rate wireless sensor networks. J. Telecommun. Electron. Comput. Eng. (JTEC) **9**(4), 37–44 (2017)
4. Vaidya B., Patel, A., Panchal, A., Mehta, R., Mehta, K., Vaghasiya, P.: Smart home automation with a unique door monitoring system for old age people using python, opencv, android

- and raspberry Pi. In: International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 82–86. IEEE, Madurai, India (2017)
5. Menezes, V., Patchava, V., Gupta, M.S.D.: Surveillance and monitoring system using Raspberry Pi and SimpleCV. In: International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 1276–1278. IEEE, Noida, India (2015)
 6. Lamine, H., Abid, H.: Remote control of a domestic equipment from an android application based on Raspberry pi card. In: 15th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), pp. 903–908. IEEE, Hammamet, Tunisia (2014)
 7. Patchava, V., Kandala, H.B., Babu, P.R.: A smart home automation technique with raspberry pi using iot. In: International Conference on Smart Sensors and Systems (IC-SSS), pp. 1–4. IEEE, Bangalore, India (2015)
 8. Sandeep, V., Gopal, K.L., Naveen, S., Amudhan, A., Kumar, L.S.: Globally accessible machine automation using Raspberry pi based on Internet of Things. In: 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1144–1147. IEEE, Kochi, India (2015)
 9. Zhang, D., Lin, S., Fu, Y., Huang, S.: The communication system between web application host computers and embedded systems based on Node. JS. In: 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), pp. 1–5. IEEE, Shanghai, China (2017)
 10. Stergiou, C., Psannis, K.E., Kim, B.-G., Gupta, B.: Secure integration of IoT and cloud computing. Future Gener. Compu. Syst. **78**, 964–975 (2018)
 11. Xue, H., Zhang, Y.: A WebRTC-based video conferencing system with screen sharing. In: 2nd IEEE International Conference on Computer and Communications (ICCC), pp. 485–489. IEEE, Chengdu, China (2016)
 12. Malhotra, J., Bakal, J.: A deterministic eviction model for removing redundancies in video corpus. Int. J. Electr. Comput. Eng. (IJECE) **8**(5), 3221–3231 (2018)

Multifactor Secure Login Using Graceful Labeled Paths



K. Vengata Krishnan and D. Lalitha

Abstract Password is a key to access secured computerized databases. Many authentication procedures exist in practice. In this article, we discuss a multifactor authentication procedure for accessing the database. Using the generic password and the gracefully labeled graph customized by a user, an authentication procedure has been established.

Keywords Graceful labeling · Multistep authentication · Array password · Password security · Password muddling

1 Introduction

In this internet age, every individual creates and maintains many social and essential accounts, each of which needs username and password for authentication purpose. Many users have difficulty in defining and remembering passwords for each of these accounts; hence, they use identical password for all their accounts. Innovations were made in creating and hashing a password. For security reasons, most of the account service providers insist on changing the password after a scheduled period of time. But users reuse their earlier password to avoid complexity in remembering a new password. This aids the professional hackers to easily hack a user account. A strong and unique password is the need of the time. The passwords that are presumed to be strong only take more attempts for the hackers, and finally, their motto is served. Jumbling a dictionary password is preferred and presumed to be secured. But a user does the jumbling that does not test his memory capacity. The vision of a researcher is toward defining a new user-friendly multistep authentication of passwords different

K. Vengata Krishnan (✉) · D. Lalitha

Department of Mathematics, Sathyabama Institute of Science and Technology,
Chennai 600119, India
e-mail: vengat0809@gmail.com

D. Lalitha
e-mail: lalkrish2007@gmail.com

from alphanumeric that are least viable for hacking. Reputed service providers in view of protecting their client information from hackers prefer multistep authentication. In multistep authentication method, service providers along with a generic password insist additional parameter for authentication. It could be a one-time password (OTP) sent to a registered mobile number or a picture confirmation which a user would have customized while creating an account. They also follow different techniques of encoding like hashing and salting, which keep their database safe from the hackers. In this article, we introduce a combination of text password and graphic password. Like a conventional method, we get a text password and then muddle them with the help of a graphical password, a graceful graph, which a user describes at the stage of registering with the service provider.

2 Literature Survey

In the existing multistep authentication methods, a generic password comprising alphanumeric characters is received as input and the account is validated to log in by verifying the one-time password, a graphical method, entering captcha, and other techniques. Even in this method, a user is required to give a very strong password which is usually difficult to remember. In 1998, Syukri [1] proposed a method of authentication by signing using mouse. Jermyn [2] proposed a new graphical technique. Abadi [3] suggested an additional parameter along with password that could be contribution by a user or computer or both. In recent past, research on array password is of big interest. The concept of array password was conceived by Vaithyasubramanian [4], following array password he gave an algorithm to create a random password using Markov chain [5]. In this method, user along with password needs to input the dimensions of an array. If the dimensions of an array are changed, the same password would be placed in different positions of an array and hence is a multistep authentication. A user has to key in his password along with the number of rows and columns of array he had chosen at the time of registering with the service provider. Vaithyasubramanian also proposed array captchas [7] for authenticating logging into a system.

3 Proposed System

In the earlier methods of multistep authentication, either the steps involve character password at each stage or the graphical password verification is independent of the character password. In this proposed two-step authentication method, a user is required to key in a character password and a graph. Based on the dimension of the graph and its label, the password is encoded and submitted to the computer. As in the conventional method of authentication, we get an alphanumeric string in the first step of authentication, and in the second step, we receive a connected graph, usually

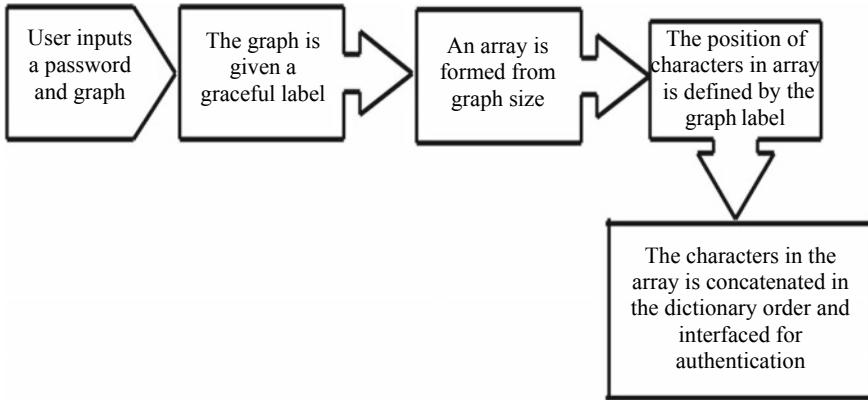


Fig. 1 Block diagram detailed the proposed two-step authentication

a path. Unlike the earlier methods wherein a user had to enter the dimensions of an array, in this process, based on the length of the path chosen by a user, we generate an array and the position in the array for each character of the alphanumeric password is decided by the graceful label given to the graph. While other graceful graphs could also replace a path, for convenience of a user who has least idea about graceful graph, we prefer a path over other graceful graphs. Once the user-defined password is arranged in an array, we concatenate it back to a string by arranging in dictionary order and we make available this modified password to the service provider's usual way of hashing. The following block diagram best explains the above technique (Fig. 1).

4 Graceful Graph

In this section, we introduce the basic definition of a graph and graceful labeling of a graph. A graph $G(V, E)$ is an ordered pair of vertices and edges of a graph. Any two vertices are adjacent if there is an edge connecting them. A number of edges that are incident on a vertex is its degree. The process of assigning a label to the vertices or edges of a graph following some rule is called graph labeling.

5 Graceful Labeling

In this type of labeling, the vertices are labeled using the labels $\{0, 1, 2, \dots, |E|\}$ or $\{1, 2, 3, \dots, |E| + 1\}$ where $|E|$ represents a number of edges in the graph, so that the induced edge labelings e_{ij} , joining v_i, v_j given by $|v_i - v_j|$, are all distinct.

In graph theory, a path P_n , is a graph with n vertices and $n - 1$ edges such that there are only two vertices of degree one and the other vertices are of degree two. Every path is graceful [6]; i.e., a path in any number of vertices accepts a graceful labeling. The graceful labeling of a path P_n is given by the following vertex labeling function:

$$f(v_i) = \begin{cases} \frac{i+1}{2}, & \text{if } i \text{ is odd} \\ n + 1 - \frac{i}{2}, & \text{if } i \text{ is even} \end{cases} \quad (1)$$

The induced edge labeling is given by $g(e_i) = n - i$.

5.1 Array Generation

We have briefly discussed in the above section the graceful labeling of a path in n vertices. Based on this labeling, we could generate an array of dimension $|v| \times |v|$ and the characters in the password that are stored in the distinct positions of array. The array positions to be filled are given by $(f(v_i), g(e_i)) \cup (f(v_i), g(e_{i-1})) \cup (g(e_i), f(v_{i-1})) \cup (g(e_i), f(v_i))$.

The following example explains the generation of array using a graceful labeling of the path P_4 (Fig. 2).

The vertices v_i are given the label from $\{1, 2, 3, 4\}$. Let $v_1 = 1; v_2 = 4; v_3 = 2; v_4 = 3$. The induced edge labels $e_{ij} = |v_i - v_j|$ are $e_1 = 3; e_2 = 2; e_3 = 1$. These edge labels are distinct, and hence, P_4 is graceful (Fig. 3).

The input password could be filled in the following array positions $(1, 3), (4, 3), (4, 2), (2, 2), (2, 1), (3, 1), (3, 4), (2, 4), (1, 2)$ in their order.

5.2 Illustration

If a user inputs “PASSWORD” as generic password and connects P_4 as graphical password, the following array would be generated:

Fig. 2 Path, P_4

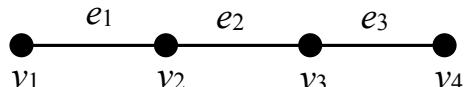
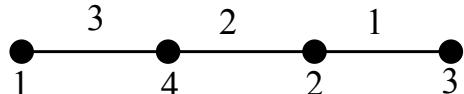


Fig. 3 Graceful labeling of path, P_4



		P	
W	S		D
O			R
	S	A	

Arranging the elements of the above-generated array in the dictionary order of their array position, we get:

P	W	S	D	O	R	S	A
---	---	---	---	---	---	---	---

Suppose for the same password a user connects P_5 as his graphical password, then the following array would have been generated:

		R		P		
	W			O		
	S				S	
D						A

Arranging the elements of the above array in dictionary order of their position, we get:

R	P	W	O	S	S	D	A
---	---	---	---	---	---	---	---

6 Conclusion

This process could be interfaced between keying in the password and hashing process. In earlier multistep authentication techniques, the user had to validate two or more independent authentication process. The process discussed above devices a new technique, wherein the final password depends on both the stages of authentication. Any normal text could be scrambled into a complicated cluster of string by the above algorithm, thereby making it difficult for the hackers to guess the actual password.

References

1. Syukri, F., Okamoto, E., Mambo, M.: A user Identification System Using Signature Written with Mouse. Springer Verlang Lecture Notes in Computer Science (1438), 403–441 (1998)

2. Jermyn, I., Mayer, A., Monroe, F., Reiter, M.K., Rubin, A.D.: The design and analysis of graphical passwords. In: Proceedings of the Eighth USENIX Security Symposium. August 23–26, 1999. USENIX Association 1–14 (1999)
3. Abadi, M., Lomas, T.M.A., Needham, R.: Strengthening Passwords. Technical Report 1997 - 033 (1997)
4. Vaithyasubramanian, S., Christy, A., Lalitha, D.: Generation of array passwords using petri net for effective network and information security. In: Advances in Intelligent Systems and Computing, Vol. 1, July, pp. 189–200. Springer India, (2014). ISSN: 2194-5357
5. Vaithyasubramanian, S., Christy, A.: A scheme to create secured random password using markov chain. In: Advances in Intelligent Systems and Computing, vol. 325, pp. 809–814. Springer, India (2015). ISSN: 2194–5357
6. Rosa, A.: On certain valuations of the vertices of a graph, Theory of Graphs (International Symposium, Rome, July 1966), Gordon and Breach, N.Y. and Dunod Paris, pp. 349–355 (1967)
7. Vaithyasubramanian, S.: Array CAPTCHAs (Completely automated public turing test to tell computer and human apart). Asia Life Sci.-The Asian Int. J. Life Sci. **1**, 247–254 (2017)

Kolam Generated by Color Petri Nets



M. I. Mary Metilda and D. Lalitha

Abstract Ordinary Petri nets and color Petri nets have been defined to generate array languages. In such models, the arrays over a given alphabet are used as tokens in the places. Using the concept of colored Petri nets and array-generating Petri nets, kolam-generating color Petri net has been introduced.

Keywords Petri net · Array token · Kolam

1 Introduction

Kolam is a traditional art which is used for decorating the temples, prayer rooms, function halls, and houses in the lower part of India. We consider the tiles used as a two-dimensional array language. Kolam can be drawn in two ways. First, we draw the dots on the floor, and we draw lines on the dots known as rangoli or we draw curves around the dots known as kambi kolam. Many methodologies have been used to create these difficult designs [1, 2].

Petri nets were first discussed by Carl Adam Petri in 1962 as a mathematical tool for modeling distributed systems [3, 4]. Tokens are used in Petri net to perform the parallel activities of the system. A Petri net has places (P) represented by circles, transitions (T) represented by bars, and directed arcs.

Colored Petri net (CPN) was developed by K. Jenson which is a high-level Petri net [3, 5]. In basic Petri nets, tokens are black dots. In colored Petri nets, data value is associated with the tokens. Conditions will be given to the transition. If the condition is satisfied, the transition is enabled. Otherwise, it is disabled.

A different kind of CPN called string-token Petri-nets was introduced. The tokens are associated with strings of symbols and the transitions with rules. On firing,

M. I. Mary Metilda (✉) · D. Lalitha

Department of Mathematics, Sathyabama Institute of Science and Technology, Chennai, India
e-mail: metilda81@gmail.com

D. Lalitha

e-mail: lalkrish2007@gmail.com

the transition takes the string from the input place and depending on the condition deposits in the output place. Array token petri-nets were presented [3, 1, 6]. In this petri-nets arrays are taken as tokens in the places. Motivated by this string and array token Petri net, this paper extends the evolutionary rule to catenate the tiles to generate kolam patterns using color Petri nets which results in a square array ($m \times m$ matrix). Each position (i, j) of the array is a tile. Some tiles may be empty tiles.

2 Preliminary Definitions

The following definitions—Petri nets, array-generating Petri nets—are recalled.

Definition 2.1 Graphically, Petri net is a bipartite-directed graph having places (represented by circles), transitions (represented by bars), and a flow relation (represented by arcs) that run from place to transition and from transition to place [4].

Definition 2.2 Array token Petri net structure is a eight-tuple $N = (P, T, I, O, \Sigma, \sigma, \mu_0, F)$, where P is a countable, non-empty set of places, T is a countable, non-empty set of transitions, I is the input function, O is the output function, Σ is the set of finite alphabet, σ is the set of rules assigned to the transition, μ_0 is the marking in the initial place, F is the final set.

Tiles used:

To generate kolam, a set of square tiles with designs in it is used. Some square tiles to generate kambi kolam are given below. The tile which does not have any pattern in it is taken as an empty tile.

All these tiles are of the same length in all four sides. We denote an empty tile by blank (without any pattern in it). Arrays represented by these tiles in Fig. 1, when joined together will create a design called kolam.

$f_1 \ l_1 \ f_2$

The array $l_2 \ c \ l_2$ taken from the tiles will form the design shown in (Fig. 2).

$f_3 \ l_1 \ f_4$

3 Catenation Rules

Let (T) be the set of catenation rules defined for every transition. For joining of tiles, we use the following four directions \implies (right) \impliedby (left) \uparrow (up) and \downarrow (down).

For example, for joining tiles d and S , we use the following rules.

$a \implies S$ joins S to the right side of a where S and a are tiles of the same length.

$a \impliedby S$ joins S to the left side a . $a \uparrow S$ joins S to the top of a . $a \downarrow S$ joins S to the bottom of a .

4 Kolam Generated by Color Petri Nets

Definition 4.1 A colored Petri net (CPN) generally defined as a seven-tuple CPNs = $\{V, \Sigma, C, N, E, \mu_0, G\}$ where V is a basic Petri net model with places, transitions, and directed arcs. Σ is a set of color set defined in CPN model, C is a color function, N is a node function, E is an arc expression function. μ_0 is the initial marking and G is the guard function.

Definition 4.2 The kolam array generated by the color Petri nets (KAGCPNs) is the set of arrays (kolam) that come to the final place P .

Example

$$\text{KAGCPNs} = (\Sigma, P, T, I, O, V, \mu_0, \sigma, F)$$

where

$$\Sigma = (\delta_1, \delta_2, \delta_3, \delta_4, f_1, f_2, f_3, f_4, d) P = (p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8)$$

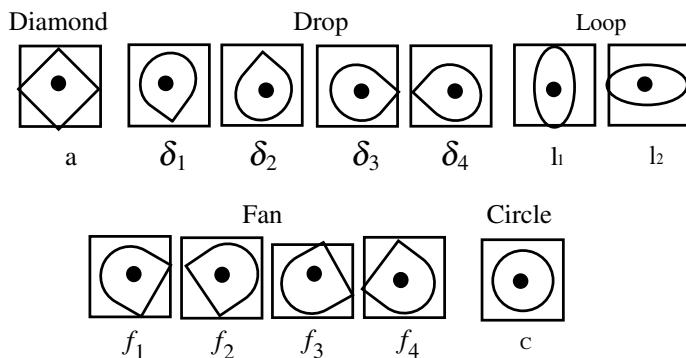


Fig. 1 Tiles used in the net

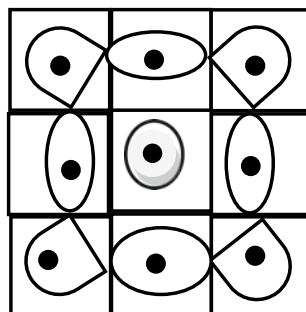


Fig. 2 Design formed by the array

$$T = (t_1, t_2, t_3), I(t_1) = (p_1, p_2, p_3, p_4), O(t_1) = (P_5),$$

$$I(t_2) = (p_2, p_3, p_4, p_5, p_6) O(t_2) = (p_7), I(t_3) = (p_2, p_3, p_4, p_6, p_7), O(t_3) = (P_8),$$

V —is the tokens used are the tiles taken from Fig. 1. μ_0 is the initial marking, σ is the set of transitions defined in the net. $F = (p_8)$ is the final set.

S is the initial array which is defined as $S = \delta_3$. i, j , and n are the integers whose initial values are also fixed as follows. $i = -n$ (for specific ‘ n ’ defined), $j = 0$ and $n = 3$ (in particular). Initial token Attribute is the set 3

$$\left\{ \begin{array}{l} < S, p_1, \delta_3 >, < n, p_2, 3 >, < i, p_3, -n >, < j, p_4, 0 >, < S_1, p_5, S_1 >, \\ < f, p_6, f_1 >, < S_2, p_7, S_2 >, < S_3, p_8, S_3 > \end{array} \right\}$$

The rules we use are defined as follows:

The rule $\delta_3 \Rightarrow a$ specifies that the right side of δ_3 is joined with ‘ a ’.

The rule $a \Rightarrow \delta_4$ specifies that δ_4 is joined to the right of ‘ a ’.

The rule $a \Downarrow \delta_2$ specifies that δ_2 is joined at the bottom of ‘ a ’. The rule $a \uparrow \delta_1$ specifies that δ_1 is joined to the top of ‘ a ’.

The firings of the diagram in Fig. 3 are described as follows

$$\sigma(t_1)\{P_1, S >, < P_2, n > < P_3, i >, < P_4, j >\} =$$

$$\{< P_1, \delta_3 \Rightarrow a >, < P_2, i + 1 >, < P_3, j >\} \text{ if } i < n - 1$$

$$\sigma(t_1)\{P_1, S >, < P_2, n > < P_3, i >, < P_4, j >\} =$$

$$\{< P_1, a \Rightarrow \delta_4 >, < P_2, -i + 1 >, < P_3, j - 1 >\} \text{ if } i = n - 1$$

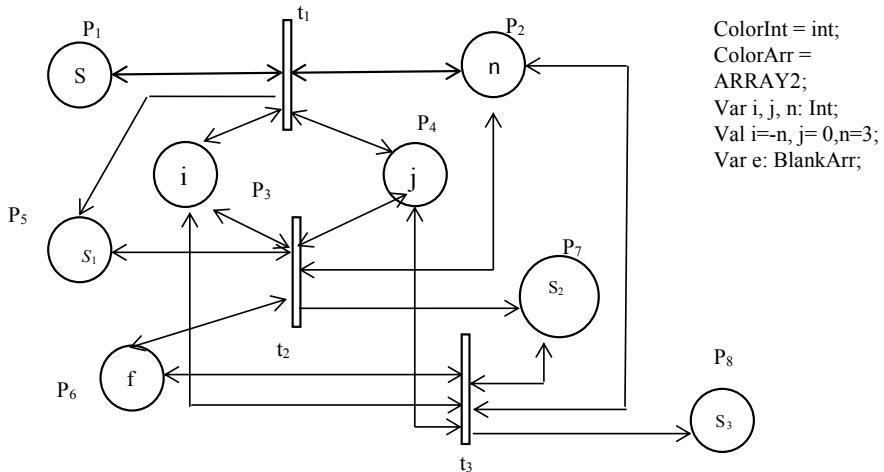


Fig. 3 Color Petri net generating kolam

$$\begin{aligned}\sigma(t_2)\{ < P_5, S >, < P_3, i >, < P_4, j >, < P_2, n > \} = \\ \{ < P_6, f_3 \Rightarrow a >, < P_2, i + 1 >, < P_3, j > \} \text{ if } i < n + j\end{aligned}$$

$$\begin{aligned}\sigma(t_2)\{ < P_5, S >, < P_3, i >, < P_4, j >, < P_2, n >, < P_6, f > \} = \\ \{ < P_5, a \Rightarrow f_4 >, < P_2, -i + 1 >, < P_3, j - 1 > \} \text{ if } i = n + j, j \neq -n\end{aligned}$$

$$\begin{aligned}\sigma(t_2)\{ < P_5, S >, < P_3, i >, < P_4, j >, < P_2, n > \} = \\ \{ P_5, d \Downarrow \delta_2 >, < P_2, -i + 1 >, < P_3, j + 1 > \text{ if } j = -n\end{aligned}$$

$$\begin{aligned}\sigma(t_3)\{ < P_7, S >, < P_3, i >, < P_4, j >, < P_2, n >, P_6, f > \} = \\ \{ < P_6, f_1 \Rightarrow r >, < P_2, i + 1 >, < P_3, j > \} \text{ if } i < n - j\end{aligned}$$

$$\begin{aligned}\sigma(t_3)\{ < P_7, S >, < P_3, i >, < P_4, j >, < P_2, n > < P_6, f > \} = \\ \{ < P_7, a \uparrow \delta_1 > \} \text{ if } j = n\end{aligned}$$

Initially, $i = -n$ (an integer) and $j = 0$, and the input place P_1 has the array δ_3 . When $\sigma(t_1)$ fires, δ_3 is removed from the input place and ‘ a ’ is joined with δ_3 in the right side till the given condition $i < n - 1$ satisfied, i will be increased by 1 ($i = -n + 1$) and j is unaltered. If $i = n - 1$, then δ_4 is joined in the right side of ‘ a ’ and $i = -i + 1, j = j + 1$, the resulting array S_1 is put in place P_5 .

When $\sigma(t_2)$ fires, an array S from P_5 is taken. ‘ a ’ is joined with f_3 in its right side and $i = i + 1$ and $j = j$, till the condition $i < n + j$ is satisfied. When $i = n + j, j \neq -n$ then f_4 is joined to the right side of ‘ a ’ and $i = -i + 1, j = j + 1$. This process will be continued till i reaches 0 and j to $-n$ ($i = 0, j = -n$). In this position $i = 0, j = -n$ δ_2 is joined in the bottom of ‘ a ’ and $i = -n + 1, j = 1$, this array is put in P_7 .

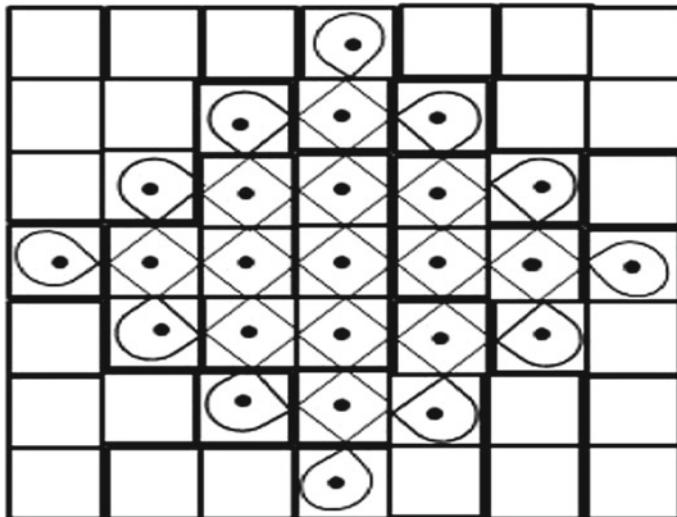
When $\sigma(t_3)$ fires, the array S from P_7 is taken, ‘ a ’ is joined to the right of f_1 till the condition $i < n - j$ is satisfied and $i = -i + 1, j = j$. When $i = n - j, j \neq n$, then f_2 is joined with ‘ a ’ in its right, $i = -i + 1, j = j + 1$. This process is continued till $i = 0, j = n$. When i reaches the position $i = 0, j = n$ δ_1 is joined on the top of ‘ a ’. The resulting array is put in P_8 . The array reaching the place P_8 is the kolam got by pasting the tiles from Fig. 1. The previous firings will generate the array in Fig. 4 the kolam reaching the final place is given in Fig. 5.

5 Conclusion

In this paper, we have considered and defined the array-generating Petri nets and kolam array token Petri nets. Tiles of the same size with pattern in it are taken as tokens to use more conditions for enabling transitions. For an easy-generating capacity, we introduced color Petri nets. Using this, we have described how a large pattern of kolams is generated in an easy manner.

$$\begin{array}{ccccccccc}
 \delta_3 & \xrightarrow{t_1} & \delta_3 & a & a & a & a & \delta_4 & \xrightarrow{t_2} & \delta_3 & a & a & a & a & a & a & \delta_4 \\
 & & & b & f_3 & a & a & a & & b & f_3 & a & a & a & f_4 & b \\
 & & & b & b & f_3 & a & f_4 & b & b & b & b & f_4 & b & b & b \\
 & & & b & b & b & \delta_2 & b & b & b & b & b & b & b & b & b
 \end{array}$$

$$\begin{array}{ccccccccc}
 b & b & b & \delta_1 & b & b & b \\
 b & b & f_1 & a & f_2 & b & b \\
 b & f_1 & a & a & a & f_2 & b \\
 \xrightarrow{t_3} & \delta_3 & a & a & a & a & a & \delta_4 \\
 b & f_3 & a & a & a & f_4 & b \\
 b & b & f_3 & a & f_4 & b & b \\
 b & b & b & \delta_2 & b & b & b
 \end{array}$$

Fig. 4 Sample derivation of the array**Fig. 5** Kolam generated by the net

References

1. Lalitha, D.: Rectangular array languages generated by a colored petri net. In: Proceedings of IEEE International Conference on Electrical, Computer and communication technologies, ICECCT (2015)
2. Subramanian, K.G.: A note on regular kolam array grammars generating right triangles. Pattern Recogn. **11**, 343–345 (1979)

3. Peterson, J.L.: A note on colored petri nets. *Inf. Process. Lett.* **11**(1), 40–43 (1980)
4. Peterson, J.L.: Petri Net Theory and the Modeling of Systems. Prentice Hall, Englewood Cliffs (1981)
5. Jensen, K.: Coloured petri nets: A high level language for system design and analysis, In: Rozenberg, G. (ed.) *Advances in Petri Nets*, vol. 483 of Lecture Notes in Computer Science, pp. 342–416. Springer-Verlag, New York, (1990)
6. Lalitha, D., Rangarajan, K., Thomas, D.G.: Rectangular arrays and petri nets, In: Baineva, R.P. et al. (eds.) *IWCIA 2012*, LNCS 7655, 2012, pp. 166–180. Springer-Verlag, Berlin, Heidelberg

Template Matching for Automatic Number Plate Recognition System with Optical Character Recognition



Arun Vaishnav and Manju Mandot

Abstract Number plate is a key feature for identifying different vehicles. The automatic number plate recognition (ANPR) system detects characters on plate on the basis of design of characters. Various font faces are used for writing and designing the number plates in India. Recognition of the characters on the number plates is subject of interest. The developed ANPR system mainly focuses on it. In this system, a new template database of English and Hindi font faces has been used for recognition and identifying of characters on number plates. The mostly used fonts are Calibri, Times New Roman and Kruti Dev for the recognition purpose in this system. The optical character recognition techniques have been employed for recognizing number plate. We have used Hindi and English character number plates for testing purpose and also analyzed the environmental conditions for the proposed system. We have found the segmentation and recognition rates around 96 and 98%, respectively, for the sample data images.

Keywords Optical character recognition (OCR) · Filter · Automatic number plate recognition · Morphological operation · Template matching

1 Introduction

Unbalanced traffic creates a lot of problems for us such as accidents, long traffic congestion, difficulties in toll collection finally resulting in delayed traveling hours, increasing pollution levels and increased fuel consumption during peak traveling hours. It is a difficult task for the parking management systems to ease the smooth

A. Vaishnav (✉)

Department of Computer Science, Mohanlal Sukhadia University, Udaipur 313001, India
e-mail: arun.vaishnav@gmail.com

M. Mandot

Department of Computer Science & IT, JRN, Rajasthan Vidhyapeeth (D) University, Udaipur (Raj) 313001, India
e-mail: manju.mandot@gmail.com

parking process during such peak hours. This has triggered a search for an efficient approach to minimize these problems. Automatic number plate recognition (ANPR) can be employed to minimize the time taken for various traffic congestion-related processes that include toll, parking collection and so on.

2 Automatic Number Place Recognition

ANPR is real-time embedded system which is used for number plate detection of the vehicles [1]. Detailed analysis of traffic, such as the number of passenger cars, public vehicles, heavy vehicles, detection of difficult traffic situation, is essential for planning new road infrastructure [2]. With these benefits, ANPR system has some basic features on which result of this system is depended like as day and night light condition, color and characters of number plate, language style of characters and location of number plate [3]. ANPR system consists of mainly three components, viz. preprocessing and identification of number plate, segmentation and recognition. These are the major steps which are used in this system for recognition of number plate [4, 5]. Qadri and Arif had explored the importance of such system and also implemented the software model using OCR [6].

ANPR uses optical character recognition (OCR), a technique of classification of alphanumeric characters from an image, for vehicle recognition. The stages of OCR include feature extraction, segmentation and classification. The other details of OCR may be found elsewhere.

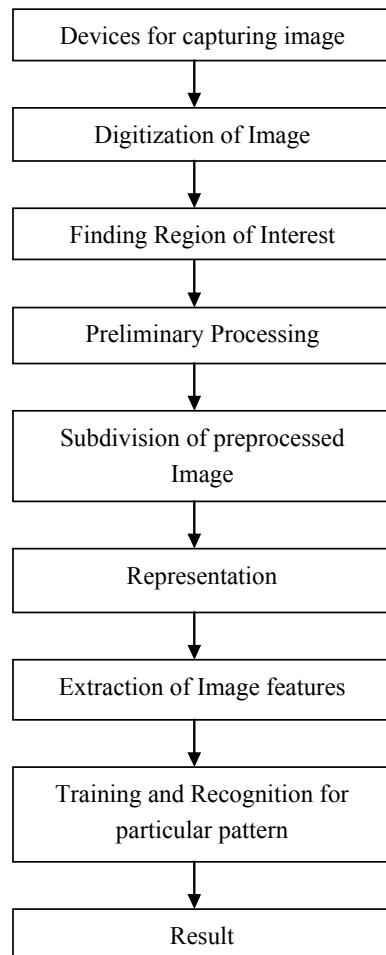
3 Automatic Number Plate Recognition with Optical Character Recognition

The concept of this technique is to teach the machine according to class of pattern. During the pattern recognition process, unknown characters are matched with previously stored character classes and find the best matches. Stored characters are trained previously in advance. The schematic representation of OCR system with its constituting elements is depicted in Fig. 1 [7].

3.1 *Digitization of Image*

Digitalization is the first step of this system by which document converts into digital form by using optical scanner. Transport mechanism and sensing devices are used for converting light intensity into gray levels. Document has two levels that are

Fig. 1 Basic elements of OCR techniques



background (white) and text (white). Such type of document is converted into bi-level by using thresholding on scanner.

3.2 *Finding Region of Interest*

The main purpose is to identify the location of an object in the image. It is determined in an image region extracted in an image according to interest of classes by using segmentation process. Each word is segmented into single characters so that it can be identified.

3.3 Preliminary Processing

Preprocessing is preliminary step of OCR system. It is used to make documents' usable stages so those documents can be analyzed. Noise removing may be done in this stage. Features of image may affect the result because it is depended on resolution, smeared characters and inherent thresholding. Recognition rate may be affected due to this reason. There are two important key points used for increasing and decreasing width of line. First is to fill the gap of broken characters for recognizing the character correctly, and second is thinning to reduce the width of line. Smoothing is used for filling and thinning. Normalization is also used in this step with smoothing to obtain certain character size and rotation. We can say that preprocessing has important objectives such as normalization, noise reduction and compression of information to be retained. Noise reduction may be performed by using three ways, i.e., morphological operations, filtering and noise modeling.

3.4 Representation

Representation is used to increase the accuracy and avoid the complexity of the algorithm. Each character is represented according to prefix characteristics or set of feature. It can be classified into three categories. Global transformation and series expansion method are used for classification purpose. A linear combination is a simple technique of representation. Transformation is a compact encoding technique of linear combination coefficient. Fourier transform, gabor, wavelets, movement are the common methods of this category. Statistical representation is the second category in which statistical distribution of points is used. This category gives benefits of speed and low complexity. Zoning, projection, and crossing and distances are the major methods of this category. Zoning is very simple method for representing the characters in which overlapping and non-overlapping zones are used. Features of different regions are analyzed for representation like contour direction features and bending features. Last category is geometrical and topological representation with style variations and high tolerance to distortions. Measuring and approximating are used for representing the characters according to ratio between height and width of bounding box of each character.

Pattern recognition is done by feature extraction. Essential characteristics of image are extracted by using this. Image feature is generally based on some characteristics or criteria which are used for evaluation, viz. deformation, noise sensitivity, ease of implementation and use. Robustness and practical usage are the important criteria. There are three categories of it, viz. transformations and series expansions, distribution of points and structural analysis. Recognition speed is the important criteria of evaluation of this method. Template matching, transformations, correlation, distribution of points are the major techniques of feature extraction. Template matching is simple and easy technique for pattern recognition.

3.5 Extraction of Image Features

Classification method is used for classifying the character in an image. There are two operations, viz. identifying characters and assigning correct class for this purpose. This approach can be divided into two categories, viz. decision-theoretic and structural methods. Decision-theoretic method is used for feature vector for representing character classification, i.e., neural network, minimum distance and statistical classifier, whereas structure method is used to derive pattern characteristics from physical structure. Relationship is used for distinguishing characters. Horizontal and vertical strokes are required for representing the characters.

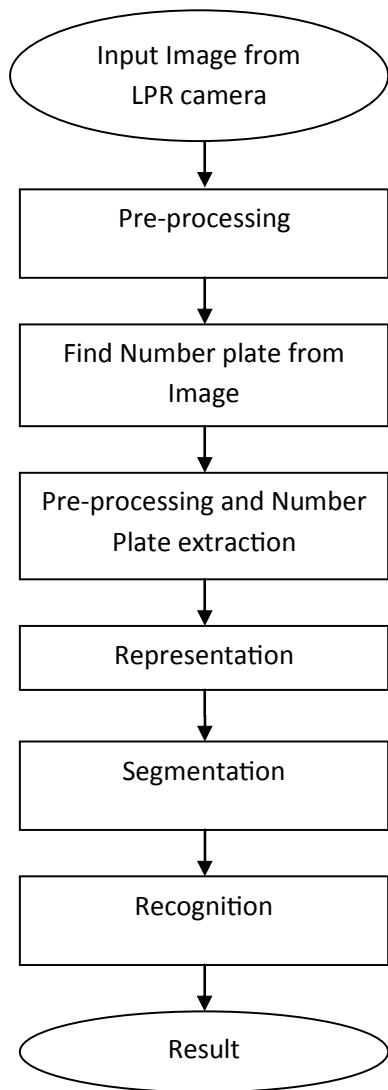
3.6 Training and Recognition

Training and recognition is a pattern recognition methodology. Pattern recognition is made up of two words, i.e., pattern and recognition. Pattern is the object, and recognition is process in which unknown object is assigned to know classes that are stored in template database. Classifiers are used for classifying the pattern. Some properties are used for representing and recognizing the patterns like curves, lines, orientation, vertical and horizontal line, and angle between the lines. For example, A having 5 lines and 0 curve that means (5, 0) is pattern for A. It is a two-dimensional space because two features are used for representing each character. In this, two terms are used, i.e., unknown sample and predefined class for pattern recognition. In this process, unknown sample is assigned into known class, i.e., known as predefined class. Major techniques of pattern recognition are template matching, artificial neural network, structural and statistical techniques. There are two approaches used by OCR in all above pattern recognition techniques for training and recognition purpose, viz. holistic and analytic strategies. Top-down approach is used in holistic strategies, whereas bottom-up approach is used in analytic for recognition purpose.

Template matching is machine vision technique of OCR for recognizing the particular pattern in an image. In this technique, there are two types of images considered. The first type of image is unknown object, and another is known object or image. Matching operation is used for identifying the correct character which is stored in template database. This operation has performed the comparison according to degree of similarity between vectors like curvature, group of pixels and shapes in the feature space.

The proposed system can also address the issue of recognition of nonstandard and irregular number plates that do not adhere to the government specifications for number plates on Indian vehicles.

Fig. 2 Flow of working model



4 Proposed Working Methodology

In this paper, we proposed an ANPR technique for number plate recognition by using camera, preprocessing by conversion, filtering, edge detection and number plate specified by morphological processing, segmentation and recognition with the help of OCR. Each step has certain functionality for achieving good recognition rate. All steps have been shown in Fig. 2.

Camera or sensors or LPR camera is used to capture the vehicle image. The first step of the ANPR system, viz. preprocessing, involves noise removal, retrieving of the characters in case of broken/damaged number plates and the images from the low-resolution camera. This is the most crucial step as any error in this step may lead to erroneous results. The various techniques that are employed for preprocessing include conversion of an image into gray and binary format, smoothing characters by using filling and thinning, normalization for obtaining same character size, filtering the image and so on [8].

Captured image is converted into grayscale image. Equation 1 is used to convert the color image into gray image [9].

$$[0.2989 * R + 0.5870 * G + 0.1140 * B] \quad (1)$$

Median is used for filtering grayscale image. It removes ‘salt-and-pepper’ noise from an image. Equation 2 shows basic about this filter [10].

$$g[m, n] = \text{median}\{[i, j], (i, j)\} \in \omega \quad (2)$$

The morphological operations and median filters are employed for noise removal, whereas the thresholding is employed for reducing the storage size. Also, appropriate thresholding results in improving the processing speed. The clear characters from number plate, filtered image and uniform shape of the characters are obtained upon completion of the appropriate thresholding. The extraction of the complete information on the number plate is done through the sequence of the following steps. Equations 3 and 4 show the morphological processing [11]. Dilation is used to repair breaks of pixels in character and erosion helps in removing the unwanted details from the image.

$$A \oplus B = \{P \in I^2 | P = a + b \text{ for some } a \in A, b \in B\} \quad (3)$$

$$A \ominus B = \{P \in I^N | P + b \in A \text{ for every } b \in B\} \quad (4)$$

In the next step, namely segmentation, each character of the number plate is divided for appropriate recognition (Figs. 6 and 7). The present approach involves explicit segmentation in conjugation with dissection for cutting up the characters of the number plate. However, it should be noted that the present method cannot be used for detection of cursive characters. Further research to improve the present method is underway and warrants a separate article. The next step in the sequence is representation of characters based on the dimension and shape for the detected characters. The classification step deals with the identification of each character and its assignment to the correct character class. In template matching, each character of the digital image is compared with the predefined database template. This step can be updated with the newer templates for English as well as Hindi font faces. An important function, namely a cross-coefficient function (CCF), is employed for

matching the similar pattern that is stored in template database. CCF is performed matching according to Eq. 5 [12].

$$\text{Correlation}(\text{image}_A, \text{image}_B) = \sum_{u,v} \text{image}_A(u, v) * \text{image}_B(u, v) \quad (5)$$

At last, result of this system has been shown in document file (Fig. 4). This result shows integration of English and Hindi characters of the number plate successfully recognized by the proposed work.

5 Analysis and Result

The proposed method can recognize the number plate which is written in English and Hindi fonts. Images are captured by camera. We have tested the method on real images; some of them are shown in Fig. 3. Recognition results of these images are shown in Fig. 4. This result is achieved by proposed methodology in which preprocessing, number plate extraction using morphological processing and recognition of license plate used by OCR template matching are used. Gray image conversion is done by using Eqs. 1 and 2 which is used for filtering. We employed Eqs. 3 and 4



Fig. 3 Images are captured by camera from straight direction. In this figure, number plate is written in (a) Hindi and (b) English

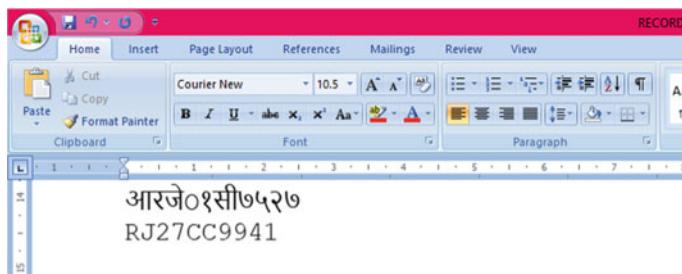


Fig. 4 Recognition results of English and Hindi character number plates



Fig. 5 Image captured from camera with different angles

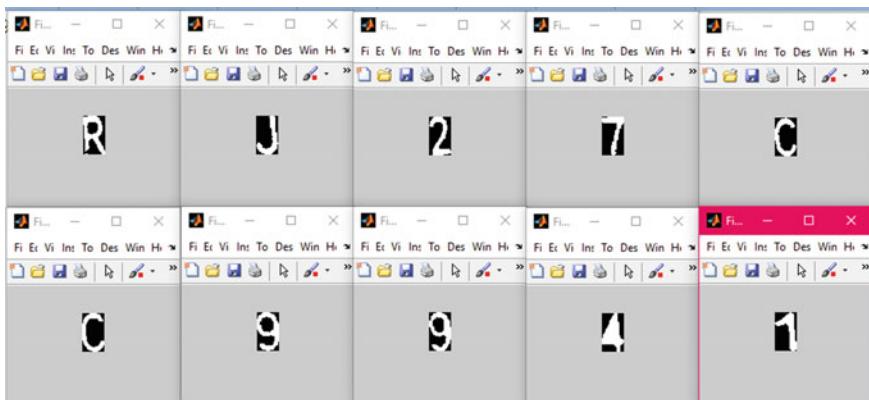


Fig. 6 Segmentation result of Fig. 5

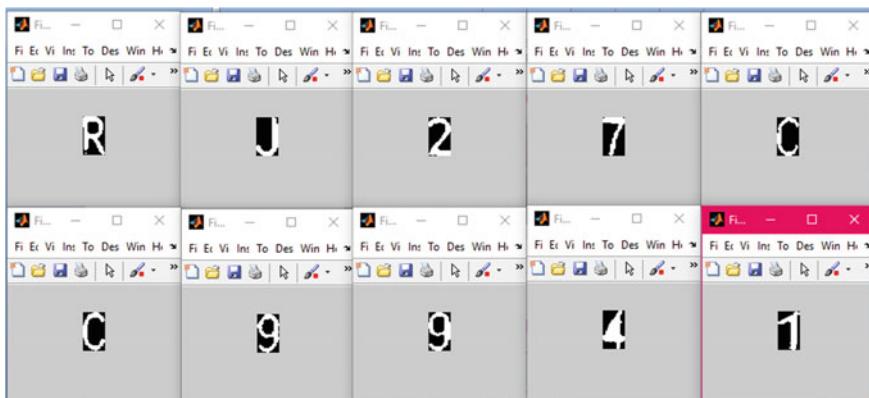


Fig. 7 Segmentation result of Fig. 3b

for morphological processing. Equation 5 is employed for finding the recognition results.

This model gives good result depending on the environmental condition as well as resource features and setting. This is also affected by angle of number plate, brightness and contrast, camera quality, lightning condition and distance between vehicle and camera.

Figure 3a shows the number plate with straight direction, and license number is written in Hindi. There are two images (Figs. 3b and Fig. 5) taken from camera with different angles. When image is taken from center front then ANPR result will be accurate (Fig. 4), whereas image captured from different angle with low brightness and contrast may affect the results. Figure 6 shows the segmentation result of Fig. 5, and recognition results are shown in Fig. 8, where character 7 in place of character 2 and 6 in place of character C are recognized wrongly. Figure 7 shows the segmentation result of Fig. 3b.

We compare the results of the proposed work with previously published work. It is indeed encouraging that the present system can yield better accuracy rate [8] with English and Hindi fonts. Also, the testing of the proposed system on the image data set of 30 images results in enhancement of accuracy rate by 2–6% (Fig. 9). We

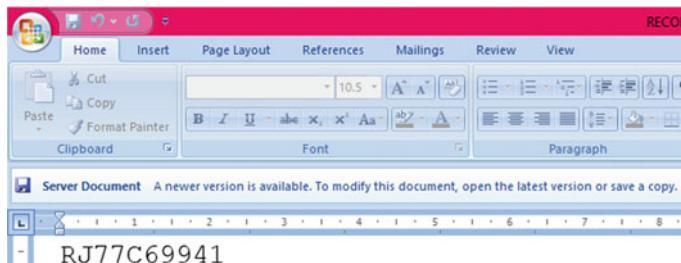


Fig. 8 Recognition result of Fig. 5

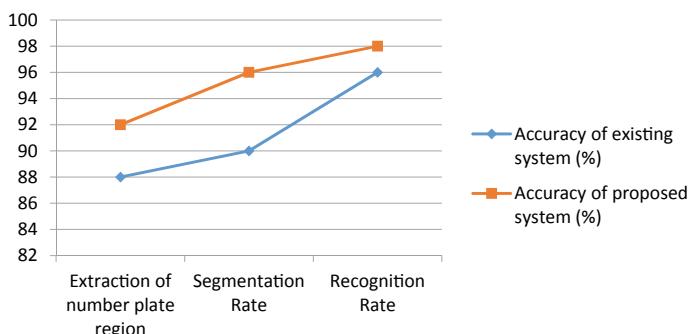


Fig. 9 Comparative analysis of the proposed work. This figure shows accuracy of the proposed work which is more than the existing system

have taken real image for testing purposes. Accuracy rate may vary according to design of license number plate. If license number plate is as per government rules or standard, then this model will be given more accurate results. The accuracy of the present approach is affected by the nonstandard number plates.

6 Conclusions and Future Scope

This paper throws the light on the importance of ANPR system for security enhancement as well as traffic management. The proposed work is based on in-house developed OCR technique that can recognize English and Hindi characters. The efficiency of recognition process to detect new template images of English and Hindi characters was enhanced by ANPR system. Also, the proposed model has been tested on a set of real images that include Hindi as well as English characters. The segmentation and recognition rates were around 96 and 98%, respectively, for the sample data set images. The extension of ANPR system to recognize more font faces and styles of other foreign and Indian languages is underway in the authors' group.

Acknowledgements The author, Arun Vaishnav, is thankful to Prof. B. L. Ahuja (College of Science, Mohanlal Sukhadia University, Udaipur) for valuable guidance and constant encouragement during the entire work. The author is also thankful to Dr. Anant D. Kulkarni (Centre for Scientific Computing, Mewar University, Gangrar, Chittorgarh) for various useful discussions and help during the preparation of this manuscript.

References

1. Kulkarni, P., Khatri, A., Banga, P., Shah, K.: Automatic Number Plate Recognition (ANPR) System for Indian Conditions. IEEE (2009)
2. Bommes, M., Fazekas, A., Volkenhoff, T., Oeser, M.: Video based intelligent transportation system-state of the art and future development. *Transp. Res. Procedia* **14**, 4495–4504 (2016)
3. Du, S., Ibrahim, M., Shehata, M., Badawy, W.: Automatic License Plate Recognition (ALPR): A State of the Art Review, pp. 1–14, IEEE (2013)
4. Sanap, P.R., Narote, S.P.: License plate recognition system for Indian vehicles. In: International Conference on Methods and Models in Science and Technology (ICM 2ST-10), pp. 130–134 (2010)
5. Anishiya, P., Joans, S.M.: Number plate recognition for indian cars using morphological dilation and erosion with the aid of ocrs. In: International Conference on Information and Network Technology, Singapore (2011)
6. Qadri, M.T., Arif, M.: Automatic number plate recognition system for vehicle identification using optical character recognition. In: 2009 International Conference on Education Technology and Computer (IEEE), pp. 335–338. Singapore (2009)
7. Chaudhuri, A., Mandaviya, K., Badelia, P., Ghosh, S.K.: Optical Character Recognition Systems for Different Languages with Soft Computing, 1st edn, pp. 9–41. Springer International Publishing AG (2017)

8. Jain, K., Chaudhary, T., Kashyap, N.: Smart Vehicle identification system using OCR. In: 3rd IEEE International Conference on “Computational Intelligence and Communication Technology” (IEEE-CICT2017), pp. 1–6 (2017)
9. Mukherjee, R.: Morphometric evaluation of preeclamptic placenta using light microscopic images. *BioMed Res. Int.* 1–9 (2014)
10. Median filter, [Online]. Available: http://fourier.eng.hmc.edu/e161/lectures/smooth_sharpen/node2.html. Last Accessed 24 Oct 2014
11. Haralick, R.M., Sternberg, S.R., Zhuang, X.: Image Analysis Using Mathematical Morphology. *IEEE Trans. Pattern Anal. Mach. Intell.* **4**, 532–550 (1987)
12. Swaroop, P., Sharma, N.: An overview of various template matching methodologies in image processing. *Int. J. Comput. Appl.* **153**, 8–14 (2016)

An Exhaustive Study of Dominant Features for Natural Gas Consumption



Prabodh Kumar Pradhan, Sunil Dhal and Nilayam Kumar Kamila

Abstract Natural gas is another type of natural resource which needs at most attention for its proper usage. Natural gas is a limited resource, and its effective usage would sustain the natural resource for a longer period of time. The proper usage and its management could be achieved through the appropriate and proper analysis and understanding of the factors affecting its usages. At the same instance, the demand and supply of natural gas need a proper strategy to meet the usage requirement. Most of the time, it is seen that the supply and demand did not match for which this natural resource is being wasted. In this paper, we made an intensive study to trace different important factors which influence the natural gas consumption. The categorization of the factors would provide a complete future strategy to effectively maintain and manage the natural gas resources.

Keywords Natural gas · Natural gas consumption · Consumption factors

1 Introduction

Natural gas [1] is a composition of methane (~95%) and hydrocarbons (~5%). These gasses are formed due to the dead marine organism's bones sunk under the submarine rocks for millions of years. During this period, the dead organisms are under high pressures and temperatures converted into gas. This gas is of two types, e.g., dry gas and wet gas. Dry gas and wet gas are mainly differentiated with methane component.

P. K. Pradhan (✉)

Regional College of Management, Bhubaneswar 69121, Odisha, India

e-mail: pkpradhan.rcm@gmail.com

S. Dhal

Sri Sri University, Cuttack 69121, Odisha, India

N. K. Kamila

Capital One, Wilmington, DE 19801, USA

e-mail: nilayam_kamila@yahoo.com

Dry gas is having more than 95% methane, whereas wet gas has less than 95% methane.

Natural gas is extracted with drilling grills, and in some areas, it is found along side with liquids and is called as associated gas. The processing of natural gas consists of the following four processes.

- a. Oil removal
- b. Water removal
- c. Natural gas liquid separation
- d. Sulfur and CO₂ removal.

This extracted gas is converted to its liquid forms and transported through pipeline to distribution centers, and sometimes, it is transported though large tankers through road ways, water ways, and railroads. This liquefied form of natural gas is called liquefied natural gas. These gasses are then further distributed for the following usages.

- a. Domestic usage for cooking and air-conditioning
- b. Commercial usage for hotels and factories, e.g., cutting irons and metals
- c. Electricity generation
- d. Vehicles and Transportation.

This is a very complex process from identifying the natural gas resources locations in oceans and sea, till its usage as mentioned above. Moreover, this natural resource is a limited resource, and its usage has to be made with appropriate forecasting and attention. In order to plan its usage, we need to make a proper analysis on different factors which affect its day-to-day usage.

This article is organized as follows. We study the different supply and demand factors for natural gas consumption in Sect. 2. In Sect. 3, we compare the natural gas consumption with the wage and weather data. We further focus on the major important affecting factors for natural gas consumption based on Sect. 3 facts and figures in Sect. 4. Section 5 concludes our work and elaborates our future scope of work.

2 Factors Affecting Natural Gas Consumption

The major factors which affect the natural gas consumption can be categorized into two categories, i.e., supply factors and demand factors. In [2] a study of the factors affecting natural gas consumption, the following are found few supply factors.

- a. Natural gas production variations
- b. Natural gas volume imports and exports
- c. Gas storage facilities.

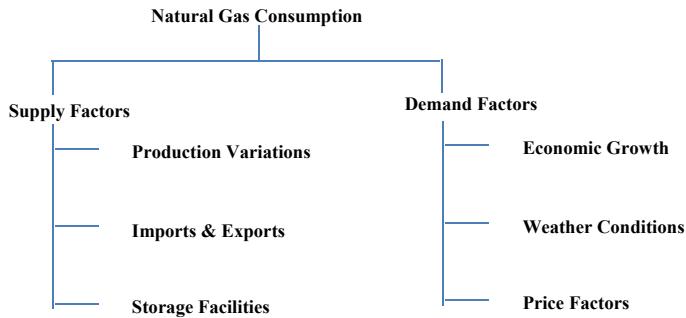


Fig. 1 Different subfactors of natural gas consumption

2.1 *Natural Gas Production Variations*

Natural gas production varies depending on various circumstances, e.g., company's production strategy and policies and variation of raw gas supply.

2.2 *Natural Gas Volume Imports and Exports*

The volumes of natural gas imports and exports are a major factor where the supply depends on. This exports and imports vary due to the import and export agreement policies, price, weather conditions, and transportation availability. Though the imports and exports are preplanned, due to the execution variations, it affects the supply factors.

2.3 *Gas Storage Facilities*

The storage facility is another important factor in supply side. If the storage is not available, then the distribution from the storage supply unit could not be achieved as per the demand (Fig. 1).

The following factors are found few factors which affect the demand side.

- Economic growth
- Weather variations
- Price factor.

2.4 Economic Growth

The nation's economic growth provides the backbone support for a better lifestyle and food style. People of the nation use the air conditions, vehicles, and quality food practices which enhance the natural gas demands.

2.5 Weather Variations

The regular day-to-day work, psychological variations of common people in food consumption, and lifestyle maintenances completely depend on the weather. A good weather can change the common people's mood to work and activities, which in turn increases the normal consumption of natural gas.

2.6 Price Factor

This factor plays a silent and important role in demand side. If price goes high even of 2–5%, it adds more expense to the monthly budget of every common people and family. If price increases, then demand will be less, and if price is less, then demand is more.

3 Natural Gas Consumption Comparison Analysis

Recent study [3] found the natural gas usage mainly depends on the developed countries and its economy status. Below table shows the country-wise data collected for top 20 natural gas consumed in 2015 in cubic meters. The following figure is the graphical representation of the nation-wise natural gas usages (Table. 1).

In [4], we collected the wage, and the following figure shows the collected data average wage per month for the respective nations.

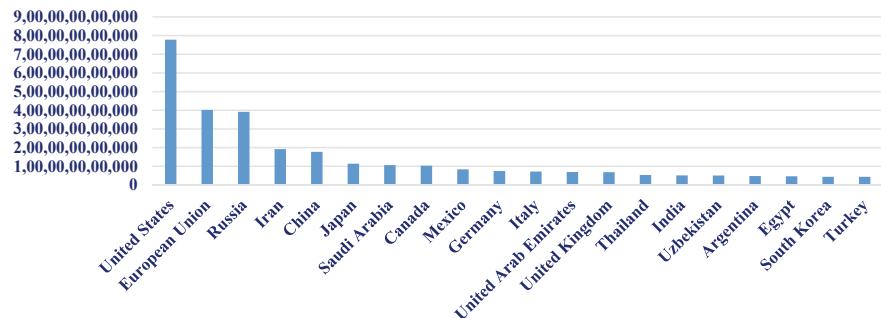
The collected data are the calculated to annual wage and then normalize as per the following equations.

$$W_c^a = \sum_{m=1}^{12} W_c^m$$

$$\|W_c^a\| = \frac{W_c^a}{\text{Max}(W_c^a)}$$

Table 1 Nation-wise usage, wage, temperature, and precipitation data

Country	Usage	Wage (\$)	Temp (F)	Precipitation (in)
United States	778,000,000,000	2949.65	52.8	34.5
European Union	402,100,000,000	NA	NA	NA
Russia	391,500,000,000	643.6	30.3	20.1
Iran	191,200,000,000	488.59	61.1	13.5
China	177,300,000,000	873.4	55.3	38.4
Japan	113,400,000,000	2509.8	57.3	67.1
Saudi Arabia	106,400,000,000	1763.24	77.6	4.9
Canada	102,500,000,000	2157.35	38	34.5
Mexico	83,200,000,000	528.66	69.1	36.8
Germany	74,600,000,000	2362.15	46.6	29.4
Italy	71,400,000,000	1479.82	56.3	30.4
United Arab Emirates	69,100,000,000	3222.34	80.6	5.1
United Kingdom	68,300,000,000	2311.89	48.8	29.7
Thailand	52,900,000,000	398.64	80.7	63.6
India	50,600,000,000	533.88	75.6	49.9
Uzbekistan	50,300,000,000	NA	NA	NA
Argentina	47,500,000,000	843.53	63.3	34.3
Egypt	46,160,000,000	156.49	71.2	2
South Korea	43,600,000,000	2113.92	53.8	51.2
Turkey	43,600,000,000	599.02	54.2	25.7

**Fig. 2** Graphical representation of natural gas consumption

where W_c^a is the annual wage for the country c , W_c^m is the monthly wage for the country c , and $\|W_c^a\|$ is the normalized annual wage for the respective country (Fig. 2).

The normalized wage is compared with the normalized natural gas consumption as shown in Fig. 3. The wage has a varied line with the country-wise consumption,



Fig. 3 Graphical representation of natural gas consumption versus wage

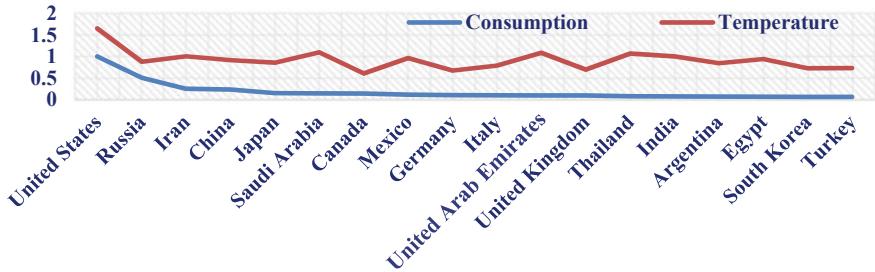


Fig. 4 Graphical representation of natural gas consumption versus temperature

but the top usage country has a overall good wage rate which enables the capability for the energy usage at its requirement.

In [5], we collected the weather data, i.e., temperature and precipitation (rainfall), for respective countries for the top natural gas usage.

We normalize the temperature data as per the below equations.

$$\|T_c^a\| = \frac{T_c^a}{\text{Max}(T_c^a)}$$

where T_c^a is the annual average temperature for the country c , and $\|T_c^a\|$ is the normalized annual average temperature for the respective country.

Figure 4 shows a comparison of natural gas consumption with temperature variations. We observed that the temperature has an approximate variation with respect to the natural gas consumption. This provides a further research area for the natural gas consumption variation relation with temperature of the local regions.

We also normalize the precipitation (rain fall) of the respective countries as per the below equation.

$$\|P_c^a\| = \frac{P_c^a}{\text{Max}(P_c^a)}$$

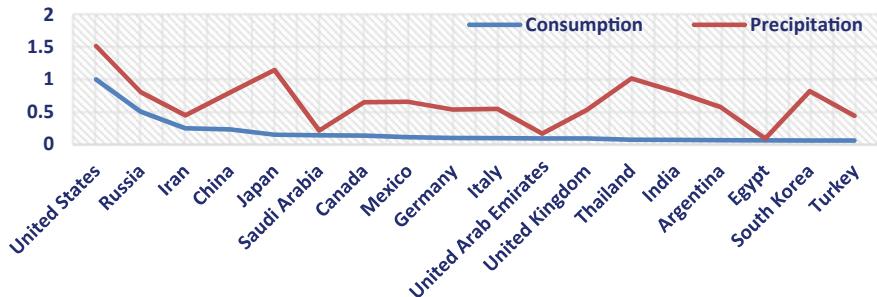


Fig. 5 Graphical representation of natural gas consumption versus precipitation

where P_c^a is the annual average precipitation for the country c, and $\|P_c^a\|$ is the normalized annual average precipitation for the respective country.

Figure 5 shows a comparison of normalized precipitation data with the country-wise normalized natural gas consumption data. We observe that the rainfall is also impacting for the natural gas consumption as it varies as per the line of natural gas consumption. There are few variations observed; however, this leads to a further research in local region rainfall variation with respect to natural gas consumption.

4 Natural Gas Consumption Analysis

We analyzed from the above facts and figures that the following factors are mainly responsible for the day-to-day natural gas consumption.

- Weather temperature and humidity
- Weather rainfall/precipitation
- Lifestyle/wage of local region.

4.1 Impact of Weather, Temperature, and Humidity

The food digestion is an inherent process of the body. The process is controlled by the body inner temperature and the external environment temperature. Our taste of food also changes based on the external environment temperature. As our food taste, style, and digestion change based on the external temperature, so we assume that the food preparation also changes depending on the external temperature. And this would be a primary reason for natural gas consumption to depend on the external temperature change. Humidity [6] on the other hand gives a different impression onto the mood and passion to consume foods. The body feels restless and worried when humidity is high. Most of the time, psychology feels restless and confused and the

mood and human behavior change according to the humidity levels. The body takes the foods based on the mood and psychology, and hence, we assume the humidity has a great impact on the food consumption and hence on the food preparation. This lead to another spike if dependency on the natural gas consumption.

4.2 Impact of Weather Rainfall

Rainfall is another major factor which impacts the supply of agriculture product. Most of the yield crops and food items are obstructed from distribution, and most of the items go waste. This creates the reduced supply to the market. As the supply to the common market is less, the preparation of cooked food items is less in variety and quantity. This correlates the less consumption of natural gas as the natural gas is used as fuel to prepare these foods.

4.3 Impact of Regional Wage

The expenditure of common people is based on the average income of the respective person's income. It is a direct equation of the common people income and expenditure. If the income varies (non-salaried people), then the expenditure varies. People of the nation cut down the expenditure based on changing the food style and consumption. If income is more or less, the consumption of food items also varies, and based on the same theory, the consumption of natural gas is also varying.

5 Conclusions

The natural gas facilitates major work, e.g., cooking and transportation in human life every day. The natural gas is a limited resource and has a very complex extraction and shipment process to enable it in daily usage. The forecasting of usage has to be made with most care and attention so that the usage life of the natural gas could be extended for a prolonged time. To make the consumption planning and future forecasting, we provided the major factors that are impacting the natural gas consumption. We provided the tabular and graphical data which in turn provide an approximated line of the dependency variations. The weather data and economic data and its graphical representation provide a deterministic relation with consumption. The forecasting model of the natural gas consumption based on the dependent factors is the future scope of this study.

References

1. Natural Gas: <https://www.studentenergy.org/topics/natural-gas>
2. US Energy Information Administration: <https://www.eia.gov/energyexplained/>
3. Pradhan, P.K., Dhal, S., Kamila, N.K.: Neural network based forecasting model for natural gas consumption. *Commun. Appl. Electron. (CAE)* (2017)
4. Numbeo: https://www.numbeo.com/cost-of-living/country_price_rankings?itemId=105
5. Weatherbase Homepage: <http://www.weatherbase.com/weather/countryall.php3>
6. Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_natural_gas_consumption
7. Ma, H., Wu, Y.: Grey predictive on natural gas consumption and production in China in web mining and web-based application. In: WMWA '09. Second Pacific-Asia Conference on. <https://doi.org/10.1109/wmwa.2009.26,2009>
8. Gilardoni, A.: Demand for Natural Gas: Trends and Drivers. In: The World Market for Natural Gas, pp 39–60. Springer Berlin Heidelberg (2008). https://doi.org/10.1007/978-3-540-68208-0_3
9. Brown, R.H., Vitullo, S.R., Corliss, G.F.: Detrending daily natural gas consumption series to improve short-term forecasts. In: Power & Energy Society General Meeting, IEEE (2015)
10. Licheng, S., Ronghua, H.: Study on the relationship between the energy consumption and economic system of Jiangsu Province base on grey relational analysis in grey systems and intelligent services, 2009. GSIS 2009. In: LIEEE International Conference. <https://doi.org/10.1109/gsis.2009.5408328> on 08 February 2010
11. Pradhan, P.K., Dhal, S., Kamila, N.K.: Time series least square forecasting analysis and evaluation for natural gas consumption. *Int. J. Recent Innov. Trends Comput. Commun.* (2017)
12. Nutrition Therapy: <http://ahnutritiontherapy.com/338/foods-to-fight-humidity>
13. EPA: <https://www.epa.gov/climate-impacts/climate-impacts-agriculture-and-food-supply>

Visual Analytics Toward Prediction of Employee Erosion Through Data Science Tools



Priyanka Nair, Jaya Krishna and Devesh Kumar Srivastava

Abstract Effective data visualization in business intelligence is an indispensable unit. The exercise of visual analytics furnishes the interactive platform to facilitate sterling user experience. With a suitable predictive algorithm to predict the outcome and visual analytics, we can generate consequential reports coalescing and combining end users' knowledge and reasoning capability. The primary tools that are used for analytics need to serve the purpose on two ends, i.e., efficient data analysis and effective data visualization. R has gained colossal popularity when it comes to its excellent data visualization potential. Moreover, with data visualization tools like Microsoft Power BI and Tableau, compelling visuals with dashboard and excellent report can be created by integration with R and incorporating R scripts and R plugins. The paper aims to scrutinize the power of visualization in predictive analytics employing rich R visuals assisted by powerful visualization tools.

Keywords Visual analytics · Data visualization · Predictive analytics · Microsoft power BI · Visual analytics using R · Tableau visualization

1 Introduction

Data visualization is the presentation and interpretation of data employing interactive interfaces and using different presentation styles (charts, graphs, maps, etc.). It provides a better understanding of data that are now no longer restricted to tabular formats. With meaningful insights generated, the use of visual analytics makes way to achieve strategic outcome and cater to better comprehension. Visualization

P. Nair (✉) · J. Krishna · D. K. Srivastava
School of SCIT, Manipal University Jaipur, Jaipur, India
e-mail: priyankanair@live.com

J. Krishna
e-mail: Jayakrishnaa.r@gmail.com

D. K. Srivastava
e-mail: deveh988@yahoo.com

is used to probe the minutiae that may not be reported or communicated in a meaningful manner with data reasoning alone. Visualization in business intelligence or visual analytics integrates three approaches that include data analysis, visualization, and individual factors (individual perception and reasoning capabilities) [1]. The human cognizance, interaction, comprehension, and analysis play a vital role in decision-making process. They play a key element in business intelligence. Visual analytics avails benefits of statistical reasoning, pattern analysis, knowledge representation, and data management. However, a considerable piece of information may go unnoticed with visualization. Even though visual analytics cannot be affirmed as an outright and standalone field of study, its role in accurate analysis of data cannot be eschewed. Visual analytics or data visualization is always backed up by statistical and analytical reasoning for better comprehension of the data set and associated patterns. Human factors alone for interpreting and analyzing the data may not be sufficient, and the comprehension requires tool support for automatically generating meaningful and useful insights to reach an astute accord [2].

The combinative impact of the exemplary statistical capability of R with interactive and user-friendly data visualization can markedly reduce the business risk by enabling the firms to delve into pattern recognition and comprehend the results. R assimilates users' knowledge and deduction potential into algorithmic and statistical data analysis processes. However, apart from graphical display of data, R does not produce insight reports. The data visualization tool Power BI provided by Microsoft serves the process of producing beautiful tabular and graphical reports which can be viewed on the dashboard and can be maintained by the organization to be able to be viewed by the organization and user and maintain the transparency. Power BI has the feature of taking the entire R script to the software and view the result on the dashboard. Power BI is a suite of business analytics tools that deliver insights throughout your organization that can connect to hundreds of data sources, simplify data preparation, and drive ad hoc analysis. It produces beautiful reports and then publishes them for your organization to consume on the Web and across mobile devices and scale across the enterprise, with governance and security built-in. The Power BI service supports viewing and interacting with visuals created with R scripts. The paper intends to present the role and importance of data visualization in the field of predictive analytics. Prediction and pattern can be comprehended with the help of visual analytics in the most efficient manner.

2 Literature Survey

Jim Thomas in his research and development plan of illuminating the path contrived the term visual analytics which is now employed in a wider context integrating different facets of statistical analysis and reasoning, data analysis and management, geo-spatial data study, temporal data processing, and human perception factors [2, 3]. A key move adopted in the field of visual analytics research was documented in John Turkey's publishing exploratory data analysis which exhibits the necessity

of shift toward exploratory data analysis [4]. Interactive visualization forms the basis for visual analytics employing statistical reasoning of data. To expedite exploratory data analysis employing visual analytics, Shneiderman et al. proposed the concept of focusing on the overview of the information first and on demand fixating zoom or filter [5]. Keim et al. extended the concept of visual exploratory analysis to gain thorough insights [6]. He explained visual data analytics as; analysis, depict the key, zoom or filter and provide the details on demand [5]. The emphasis on human and computer association in data visualization process was brought to attention by Crouser et al. [7]. Relevance and usefulness of human interaction in data visualization process were submitted by Simoff et al. [8]. Munzner et al. proposed four layers of data visualization process design that includes domain problem characterization, data abstraction design, interaction technique design, and algorithm design [9].

Various data mining and machine learning techniques may be adopted to study and analyze the data automatically. Numerical and algorithmic approaches have proven to be useful for deducing meaningful insights pertaining to numerous roadblocks that include limitation like expanding data size and algorithm scalability. Users have a key role in analyzing the data by providing their knowledge and reason the data under different analysis scenarios.

3 Research Methodology

The visual analytics process is a combination of automatic and visual analysis methods combining with human interaction and efficient data analysis in order to gain knowledge from data. The analysis of data visualization efficiency has been carried out on 23,326 rows of company data that include employee demographic and job characteristics with predictive algorithm processed through the data. The solution overview to predict the employee churn and visualize the report graphically is shown in Fig. 1.

The objective is to visualize and identify the high-risk cluster of employees so as to take appropriate steps to retain the talent. To predict the employee churn, we will transform the raw employee data by applying the data transformation methods of missing value and outlier treatment, data normalization, data scaling, treatment of categorical variables, and establishing correlation. On the transformed data set, predictive modeling algorithm is applied to determine the predictive statistics pertaining to the employees who are susceptible to leave the organization in immediate future. The predictive algorithm taken herewith is decision tree modeling. The modeling can be carried by employing a data modeling tool R.

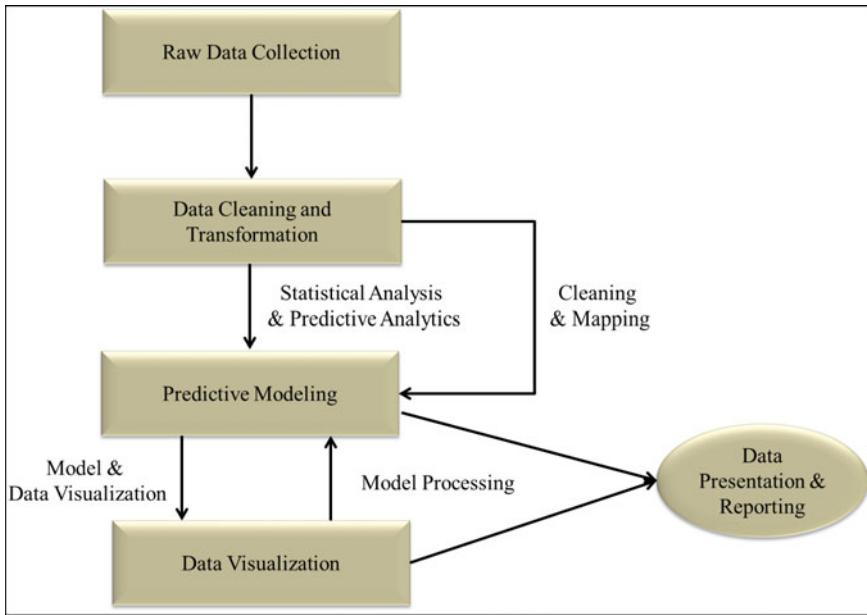


Fig. 1 Solution overview of visual analytics in predictive analytics

3.1 R Data Visualization

The employee data set of 23,326 rows of data comprises of 19 data points with 17 predictor variables and 1 target variable. The target variable is The HR.STATUS which implies the status of employees as active or inactive. With the transformed data set, we will consider only nine predictor variables. Classification and regression trees (CARTs) can be generated through the rpart package. Decision tree is a graph to represent choices and their results in form of a tree. The code for rpart decision tree formula is shown below.

```

library(rpart)
#grow tree
Fit ← rpart (HR.STATUS ~ AGE + GENDER + MARITAL.STATUS +
DEGREE.DESC + NO.OF.PRIOR.EMPLOYERS + ATTRITION.REASON +
TOTAL.EXPERIENCE + LEAVER.TAKEN + DEPARTMENT, method=
“anova”, data = trainDF)
  
```

Plot the tree to view the splitter nodes, and determine the data points associated with attrition. The code for plotting tree is given below.

```

# plot tree
plot(fit, uniform = TRUE, main = “Decision Tree for predicting Employee
Attrition”
  
```

```
text(fit,use.n = TRUE, all = TRUE, cex = .8)
```

However, the plot shown with the above code is not easily readable and cannot be comprehended easily; hence, we accustom visualization feature of rattle package of R. Rattle is an open-source GUI for data mining using R, and it employs Gnome GUI. Rattle brings together a congregation of R, and it adds sophistication in terms of visualization to our data. Rattle include ellipse, gplots, randomForest, reshape, rggobi, RGtk2, mice, network, party, rpart, and many other packages that are all accessible from the comprehensive R archive network (CRAN) [10]. For our current work, rpart package has been used. We here employ fancyRpartPlot. It has the following syntax: fancyRpartPlot(model, main = "", sub, caption, palettes, type = 2, ...). The rattle package code is given below to present the result in a more readable manner (Fig. 2).

```
Install.packages('rattle')
library(rattle)
library(rpart.plot)
library(RColorBrewer)
fancyRpartPlot(fit)
```

Observation: The data visualization component of R makes it easier to interpret the data 3 data points of 9 predictor variables have been employed to build the tree. The employees with their marital status to domestic and household head are at minimal risk of quitting the job and are expected to be stable and continue the job. Ten percentage of the employees is likely to stay. The employees with the marital status to be single, married, divorced, widowed, and separated may or may not quit the job with further splitting with the variable total experience. The employees with a total experience of more than 9 years are at highest risk of attrition. The employees with less than 9 years of experience may again be expected to be stable or unstable. The employees with age greater than 31 are more stable than those employees with age lesser than 31. of reusable components which is a regulated consolidation of altered and new source code.

3.2 Tableau Visualization

Tableau supports advanced analytics that is capable of unveiling the complexity of the problem, spot and share visualized predictions. Tableau is one of the rapidly enrolling data visualization tool. It incorporates powerful input capabilities, drag-and-drop facet, and dashboard creation that uniquely distinguish it from other products. The interactive and easy to understand dashboard facilitates even the inexperienced professionals to bring their data to life [11]. With this BI tool, advanced visualization is expedited which encompasses data representations in the form of charts like Pareto, box plots, funnel charts, and histograms. It also bolsters segmentation and cohort analysis. It paves way for time series and predictive analytics. Tableau is also well

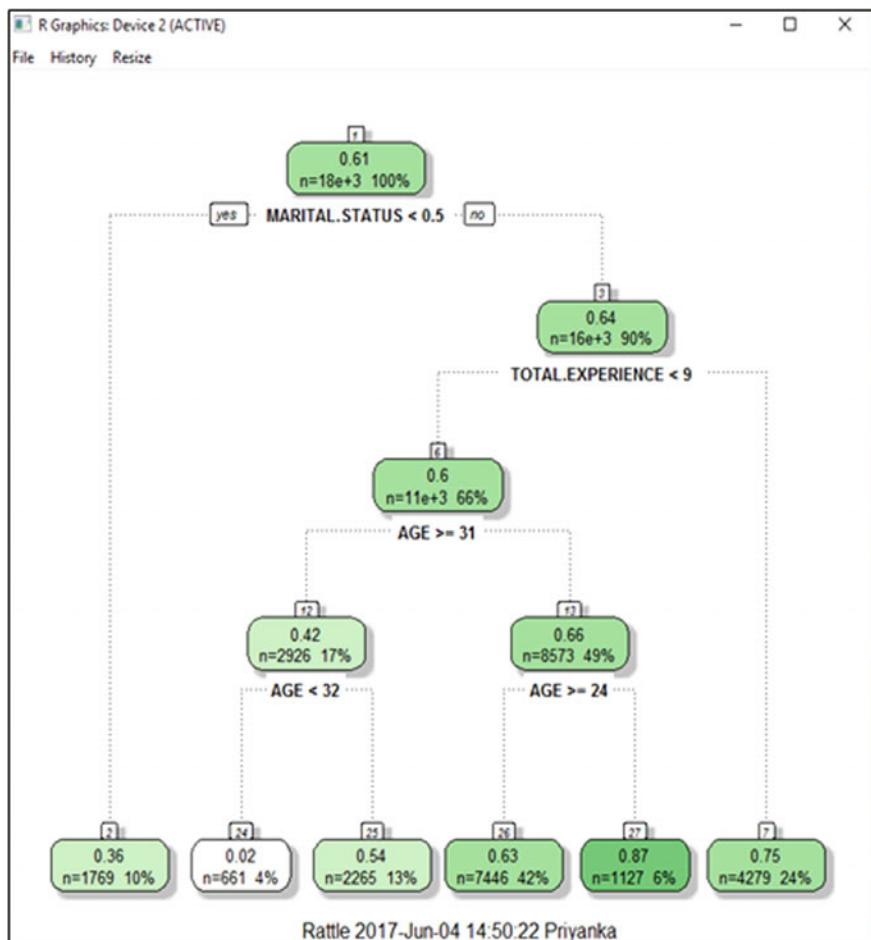


Fig. 2 Variants of decision tree plot with and without dummy code

suites with integration of external services like R plugins. Statistical modeling and analysis with R can be leveraged alongside the front-end capabilities of Tableau to generate intriguing and meaningful insights with data. Tableau facilitates advanced data mining and depicts statistical results in a comprehensive manner.

For the same data set, if we run an ad hoc AB-Test with dependent variable HR.STATUS and independent variable AGE, we can efficiently visualize the age distribution with the percentage in each category with the employment status as active and inactive, respectively. Figure 3 depicts the age distribution generated in Tableau on record set of 10,000 rows on account of employment status as of active or inactive. Figure 4 shows a more consistent distribution of age with status marked with a reference line.

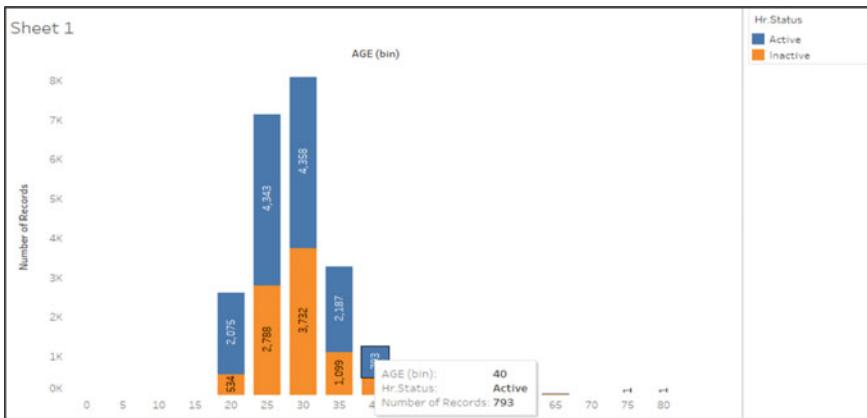


Fig. 3 Age bins and employment status of employees with respect to age

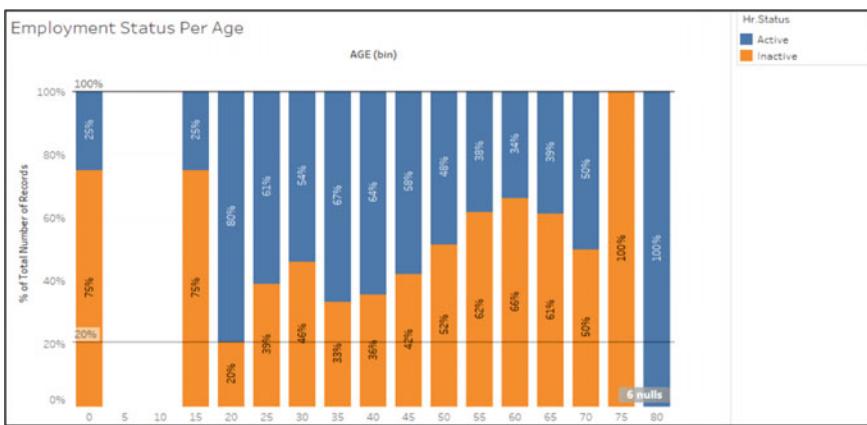


Fig. 4 Consistent employment status with a reference line

3.3 Power BI Visualization

Power BI is a cloud-based application that lets one hand-pick from a catalogue of available and obtainable visualizations generate custom or stipulate visualization from community gallery. R visualization, however, does not generate reports and additional insights pertaining to the data points available with the data set. Power BI facilitates the incorporation of Rscript to view the results on the dashboard element of BI desktop and generate useful results to be maintained by the organization. Power BI has rich question and answer facet that changes how an entire firm ruminates and interacts with the data. Business analysts integrate data analysis and rich visualization to build substantial reports and scrutinize the key business bits. Powerful analytical tools caps data to insight where data can be combined from multiple sources and

High Risk Cluster				
HR.STATUS	Masked Employee ID	TOTAL EXPERIENCE	DEPARTMENT	
A	U100038	14.42	Delivery - Healthcare	
A	U100533	16.20	Delivery - Anthem	
A	U100804	10.70	Delivery - Anthem	
A	U100837	16.57	Delivery - SemiConductor	
A	U101040	10.25	Emerging Services- IoT	
A	U101242	10.16	Infraserv-Cloud Infrastruc Sol	
A	U101288	9.61	Delivery - HiTech	
A	U101491	10.51	Delivery - Banking & Financial	
A	U101623	10.92	Accounting and Compliance	
A	U101692	9.75	Delivery - Emerging Accounts	
A	U101797	10.05	Delivery - Europe	
A	U102086	15.36	EnggSer-Mechatronic	
A	U102137	10.20	G&A - Equity Investments	
A	U102210	9.14	Delivery - Talent Pool	
A	U102349	20.77	G&A - Europe	
A	U102739	11.52	Emrg Serv-Cloud	
A	U103106	9.34	G&A - Maersk	
A	U103137	11.19	Delivery - Europe	
A	U103154	13.72	Delivery - SemiConductor	
A	U103788	15.54	Delivery - Europe	
A	U103949	10.93	Sales - Business Services	
A	U104084	16.62	Delivery - Banking & Financial	
A	U104228	10.21	Delivery - Banking & Financial	
A	U104279	11.96	BSE	
A	U104864	32.43	InfraServ-SpecialisedApplnSolns	
A	U105175	12.49	Delivery - Banking & Financial	

Fig. 5 High-risk employee cluster generated with Power BI

vigorous and rich analysis can be carried out to bring their data to existence. BI provides unique experiences for business analysts to work with data. BI Desktop is incorporated with drag-and-drop data features and interactive and customized reporting.

The decision tree algorithm processed with Rscript is taken entirely to desktop and the rule set is generated automatically with the tool. With Power BI Desktop, one can perform analytical and statistical analysis and create compelling visuals by integrating with R. One can host those R visualizations within the Power BI Desktop report. For the data set used, Fig. 5 shows the high-risk employee cluster who can

take an exit in the immediate future. Such reports can be generated with Power BI tool to assist the management with specific data, and employees can be hand-picked for further counseling to maintain the retaining scale.

4 Conclusion

Data visualization cannot be considered as a standalone segment; it can generate meaningful insights to the data with tight coupling with human perception and statistical reasoning. The current work explores the visual element of R which can generate graphics pertaining to summary statistics generated. For a novice employing R and features of R, visualization lays challenges, but the visualization tools like power BI and Tableau can be efficiently used by non-experts to transform and analyze their data. Rscripts can be incorporated with the Microsoft Power BI tool, and plugins can be used with Tableau to view the personalized result set in a dashboard format. Power BI generates reports that can be hosted online and can present the data graphically and gives an interactive environment to the user. Tableau can generate powerful insights and perform clustering to generate deducible facts in chunks. The power of visual analytics has been demonstrated in the field of predictive analytics by taking up decision tree predictive modeling to predict the employee cluster at the risk of leaving.

References

1. Eckerson, W., Hammond, M.: Visual Reporting (2011)
2. Keim, D.A., Mansmann, F., Thomas, J.: Visual Analytics: Scope and Challenges. Tutorial at IEEE Visualization (2007). <http://infovis.uni-konstanz.de/tutorials/>
3. Thomas, J., Cook, K.: Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press (2005)
4. Tukey, J.: Exploratory Data Analysis. Addison-Wesley, Reading MA (1977)
5. Shneiderman, B.: Eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the IEEE Symposium on Visual Languages, pp. 336–343 (1996)
6. Keim, D.A., Mansmann, F., Schneidewind, J., Ziegler H.: Challenges in visual data analysis. In: Proceedings of the IEEE Conference on Information Visualization Oct 2006, pp. 9–16 (2006)
7. Crouser, R.J., Chang, R.: An affordance-based framework for human computation and human-computer collaboration. IEEE Trans. Visual Comput. Graphics **18**(12), 2859–2868 (2012)
8. Simoff, S.J., B’ohlen, M.H., Mazeika, A.: Visual Data Mining. Springer (2008)
9. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Visual Comput. Graph. **15**(6), 921–928 (2009)
10. Williams, G.J.: Rattle: a data mining GUI for R. R J. **1**(2), 45–55 (2009)
11. Wesley, R., Eldridge, M., Terlecki, P.T.: An analytic data engine for visualization in tableau. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data, pp. 1185–1194 ACM, (2011)

Secured Forensic Framework for Various Users in the Virtualized Environment of Cloud



Gayatri S. Pandi and K. H. Wandra

Abstract A continuous forensic system, made available on the Cloud Server, which is built and integrated with the architecture of the cloud will help in gathering expedient additional “evidences” against certain crimes which could be useful in framing the criminals in the court of Law in any country like, e.g., India, when cloud services are used to commit the crime. The evidences have to fit into the system which are complaint with the policies set by the Law regulatory authorities of India. Most of the software agents generate Log data and store them in an Insistent Database which is centralized by nature. Such centralized systems can be manipulated by certain malicious attackers who can even be the employees of the organizations or the organizations itself to prevent penalties for non-compliance to standard policies if such systems come to notice when a crime is being committed. The manipulation can be to such an extent that the evidences can be modified or even swapped with possibly false Logs to frame some innocent users. This research paper proposes a cloud-based framework termed as Forensics Braced Cloud for gathering of expedient evidences and keeping such crucial expedient evidence Logs in a confidential way on another server not under the control of Cloud service provider. The Logs kept on the Insistent Database are used to publish the evidences and can be later checked by the Law regulatory officer when needed to check the integrity of the evidence termed as evidence of past Logs. The results observed are currently specific to the Prototype environment.

Keywords Cloud forensic system · Insistent database · Evidence for past logs · Authenticity · Integrity

G. S. Pandi (✉) · K. H. Wandra

L.J.Institute of Engineering and Technology, Ahmedabad, India

e-mail: gayatri.jain@ljinstitutes.edu.in; gayatree.jain@gmail.com

1 Introduction

A continuous forensic system is the real need in today's world where data is distributed on many servers and so can be the crime for attacks like distributed denial of service. NIST defines Cloud forensics as “the application of scientific principles, technological practices and derived and proven methods to reconstruct past cloud computing events through identification, collection, preservation, examination, interpretation, and reporting of digital evidence” [1, 22]. Current Cloud architectures lack great support for digital forensic investigations and are not compliant with the digital forensic procedures as needed in the present-day world due to its dynamic nature as discussed by Alqahtany [2]. To track and prove the crime legally in a country like India, one needs to apply forensic technology. Cybercrime is evolving at a very fast pace, and the criminals apply a remarkable effort to derive major benefits from the vulnerabilities observed in the architectures as discussed by Marturana [3]. There is a lack of forensic readiness in Cloud computing policy framework as discussed by Spyridopoulos and Katos [4]. It is a great challenge to apply forensic technology in Cloud computing environment firstly due to the distributed nature of the cloud data and the processing. Secondly in Cloud computing environment, the forensic investigators do not have an appropriate control on the evidence as it is in the traditional computing environment. In the later environment, it is easy for the legal authorized forensic investigators to seize the computer and the other evidences in their own custody in case of suspicion of crime being committed using them, but so is not the case in Cloud computing environment where the entire data center cannot be kept in the custody as the data leading to the evidences can be distributed to many such centers. On the third point, it is a challenge to collect the reliable evidence due to its multi-tenant features. Fourthly, when the virtual machines are shut down may be intentionally or otherwise leading to loss to data and it becomes difficult to store volatile data. Once the volatile data of the virtual machine is lost, the crucial expedient evidence of who was allotted those virtual machines and what all websites they accessed or what all they did using those hired virtual machines is lost. So, it becomes a great challenge to handle such crimes and frame the criminals. On the fifth note, the forensic investigators are completely reliant on the Cloud service providers (CSP) for acquiring the evidences. Also, the confidentiality and the integrity of the evidence is a biggest challenge. Finally, the custody of the evidence, the way it is gathered, analyzed, and preserved also remains a big challenge in many countries. Maintaining integrity of such evidence remains a biggest challenge. There is a tangible need to prove that in such evidences Integrity needs to be proved mathematically.

2 Related Work

Cloud forensics is one of the challenging domains of research and relatively new. A lot many issues observed and challenges faced are discussed by Gayatri et al. [5]. As

the cloud is distributed in nature and effectively uses the underlying network which could be private or public. Cloud forensics is defined as the subcategory of network forensics by Ruan et al. [6, 27]. They described three dimensions of Cloud forensics, i.e., the technical, organizational and legal dimensions. This kind of definition is incomplete as investigating the hard disk stored in Amazon's infrastructure is not just network forensics.

Diane Barrett has discussed about the type of evidences that can be assembled in virtualized environment for forensic investigation, and the architecture has to be secured [7]. Patrascu et al. built a Cloud architecture to monitor the activities. They suggested that the Logs need to be gathered from different layers of cloud. The authors also implemented a data center topology on which the proposed framework was executed [8]. Indrajit Ray et al. proposed a framework with a cryptographic protocol to address the integrity and confidentiality and other issues that were integrated in the host operating system in the cloud and computers and the servers are secured through the use of anonymizing the network [29]. Mustafa et al. have argued about issues in cloud forensics such a confidentiality, performance, and network overhead and integrity [9]. The authors discussed about the issues faced in instance image acquisition and the integrity for both the object and block storage is verified. Sean Thorpe et al. have conversed about various service delivery models for the cloud with intermediate cloud brokers with diverse SLA agreements. The forensic analysis done must be appropriately dispatched to the end users [10]. Trenwith et al. applied hash algorithms which generate a hash log which is employed like an encryption key to encrypt a salt value, and the output of this process the cipher text is saved. It employs the encryption key to encrypt the salt value. Also additions are made to the metadata system [11]. To certify the integrity of the evidence, Shams et al. [12] recommended a digital signature process can be applied to the data collected as evidence and a procedure needs to be defined to check the same later. The researcher has created Log chain and used RSA algorithm heavily for encryption and SHA-256 algorithm for hashing. One-way accumulator scheme is designed using RSA algorithm [13]. This algorithm has its own limitations of key size, and cryptographic strength when higher level of security is essential [14]. Bloom filter and RSA accumulator have been used to build the proofs of past logs. Another accumulator scheme named bloom tree is applied which the researcher claims to accomplish better results when compared to the other two accumulators in terms of time and space requirement [15, 16].

FROST a forensic data collection for OpenStack by Dykstra et al. was developed for the cloud users and the investigators. It was employed to get the image of the virtual disks linked with any of the user's VM, and one could also check the integrity of those machines with cryptographic checksums [17]. They also supported in providing logs for all the requests made to the CSP and OpenStack logs of the VMs. But here, the focus was to just make the logs available to the different users. They also relied a lot on the CSP or rather assumed that the CSP is honest. Cloud logs in Cloud environment play a significant role for evidence when crime is committed using cloud. These logs need to be secured as they can be modified by the criminals or any other entity. As a precaution and the need to secure it from criminals, the logs need to be secured as discussed by Sang [18]. Marangos et al. introduced a digital

forensic hypermodel which describes the basic steps applicable in the digital forensic world [19]. As the technology progresses, the log collection techniques need to keep pace with it. It is very much required to build framework of methodologies and process for cloud forensics that will aid in building reliable evidences acceptable in the court of Law in countries like India against the cyber criminals.

Our contributions include the following:

- (1) We introduce a continuous forensic Logs system to monitor and collect the Logs which can be termed as evidence created by the users of our system. Our proposed system preserves the confidentiality of the users who are termed as clients to the working system.
- (2) We introduce the collection of evidence for past logs (EPL) which cannot be altered by any of the malicious entities like Cloud service provider or any employee is the organization or even the forensic investigator.
- (3) The proposed system proves the confidentiality and the integrity of the evidences gathered using best key generation encrypting algorithm and the best hashing algorithm as per the need of the Law of most of the countries including India.
- (4) The proposed system is built on top of open source Cloud computing platform—OpenStack. It is possible to integrate the module built using Python code as a part of OS or as a layer of service in the future and solves the issues of volatile logs of virtual machines.

The sections are organized in the following manner: Sect. 3 specifies about the security requirements for the Cloud log forensics, Sect. 4 focuses on the proposed work, Sect. 5 discusses about the threat model followed in this paper, Sect. 6 explains about the prototype implementation and results observed, and Sect. 7 concludes this research paper with some inputs on future work.

3 Security Requirements for the Cloud Log Forensics

A secured forensic framework is a need in today's world which aids in gathering expedient additional "evidences" against certain crimes which could be useful in framing the criminals in the court of Law. Security and Forensics are compulsorily needed together as the better understanding and usage of security is a precondition for better indulgence and application of Forensics as discussed by Simon [20]. The needs and features of such framework include confidentiality, integrity, availability, authenticity, and privacy.

Confidentiality can be achieved by conserving all client's data in the Log files and protecting it against the manipulation by some attacker. The attackers in this scenario could be any employee of the CSP's organization or any competitors of the CSP. When such client's Log files are scrutinized, it is a matter of concern as it has numerous client's data put together at the same time and can be misused by anybody. If a specific client's data is accessed, it can lead to retrieving some other client's data as well in parallel. An investigator can also retrieve the client's data

during investigation and can make alterations intentionally or even by accidents. Such issues will be leading to compromise on the confidentiality aspects. The CSP must ensure such that the level of trustworthiness is not at stake or exploited and all the Client's information is safe and not manipulated.

Integrity of the log files is very essential as they are used as evidences in the court of Law against the attackers and a lot of effort is needed to maintain it. A lot of research is focused on this domain of evidence integrity. Tampering or modification of evidence in the Log files is possible if a crime is committed using the Cloud virtual machines and if the Logs provided by the CSP are not secured against such malicious activities. Such events could lead to big issues like an attacker can commit a crime and hide himself thus encouraging him to commit many such crimes. Such activities must be curbed by ensuring the Logs are secure and the Integrity is not compromised. The CSP has to provide the original Logs and the forensic investigator has to prove in the court of Law that the integrity of the Log files is intact.

Privacy is defined as to shield the client's data from being disclosed to unwanted personages. The Logs go through the different phases right from the generating phase to investigation phase. As multi-tenancy is a major significant characteristic of Cloud computing environment where one physical resource may have many virtual machines created on it and many clients' applications are hosted on these virtual machines simultaneously leading to resource pooling environment and also leading to generation of Logs clubbing many clients' data together. Such issues are discussed by researchers like Simou et al. and Jahlali et al. [6, 21, 26]. The authorized personages should be accessing only the attacker's Logs and not the other client's data who are innocent or hold no hand in crime. If, by chance, the innocent client's data is accessed, it could lead to issues of violations of privacy. Thus, security features like privacy play a major role and remain a major challenge to keep track of, for the forensic investigators.

Availability is defined as the data must be accessible as and when required in the original form. There is a need to create multiple replicas of the Cloud Log files at different places for the purpose of security. The Logs also need to be consistently reliable. Availability of Cloud Log files is very significant as they become the source for the investigators who search for evidences in these Logs. But it will be only helpful if the data is complete and not attacked that is it should be accurate.

Authenticity means only the appropriate authorized personages are given permission to access the Logs. The authorized personages can be an investigator who is scrutinizing such Logs to trace the evidences. The Logs may be accidentally and intentionally deleted or amended by the investigator or any authorized or non-authorized employee of the CSP. If a crime is detected in these logs and needs to be proved in the court of Law, it becomes very essential for the CSP to justify that such Logs are accessed by personages who have legal permissions. Also, a Log needs to be maintained about who has accessed these Logs and when. Incase these Logs are accessed by the non-authorized persons, the Logs can be disclosed leading to unwanted issues of security breach which could cost heavily to the CSP in terms of business as many clients may stop using the services provided by the CSP. Thus, authenticity plays a significant role and is an important feature of security.

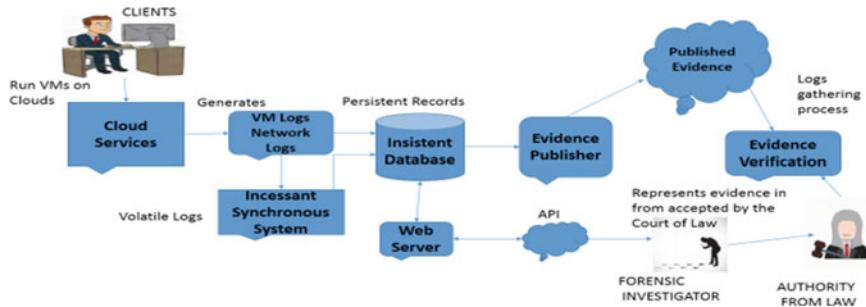


Fig. 1 Forensics braced cloud

Step : 1	Step : 2	Step : 3	Step : 4	Step : 5	Step : 6
Preservation	Incident		Examination		
	Identification	Collection	Organization	Presentation	Verification
	Evidence		Analysis		

Fig. 2 Cloud forensic process flow

4 Proposed Work

In our proposed system design, we design a Forensics Braced Cloud (FBC) as shown in Fig. 1. A digital forensic process (DFP) goes through the various stages like the identification, collection, organization, and presentation. There is an essential need for the introduction of preservation and verification in the existing process of DFP for the Cloud forensic process (CFP) as the cloud data is volatile in nature and many manipulating entities can exist like the CSP or an employee working at the CSP or the forensic investigator as shown in Fig. 2. Thus, the preservation stage has to be online termed as Insistent Database which is continuously throwing data on the web server as shown in Fig. 1.

5 Threat Model

Let C denote a set of all clients who are the users for the set of services S (any type of service hardware, software) provided by the CSP. There could be many clients say m clients and many services say n services provided by the CSP. A set of virtual machines can be used by the clients denoted as V . Virtual machine usage generates a set of logs of the services used by the clients denoted as L . These Logs can be used as a set of evidences in the court of Law in case some cybercrime is committed. A forensic investigator can perform the analysis when a cybercrime is committed using these set of services and presents the set of evidences in the court of Law. Mathematical notation for all this is provided in Table 1.

A client C who ever uses the services S hosted by the CSP by using the virtual machines V , the system has to generate the Logs L which will be the evidence for past logs. Our proposed system creates all the evidences for client's Logs, which is represented as the union of all the Logs generated for the respective clients as shown below:

$$E = S \cup V \cup L \cup C \quad (1)$$

A hypothetical case study of Buzz coffee by Dykstra et al. [22] also discussed by Suleman et al. [23] where a malicious user installs a rootkit to remain undetected by any software after performing the planting attacks. The malicious user victimizes the Buzz coffee company's website installs a malware. The innocent user's network traffic is directed toward the web page with malicious content, and the site performs a malware attack on them. Such cases remain a challenge for the investigators. Due to lack of support from the CSP, the case remains unsolved. We need to depend on the CSP for logs while performing investigations as discussed by Patrascu et al. [21]. It is otherwise very difficult to gather the required data as discussed by Alqahtany et al. [24].

In our implemented system, the CSP has to deploy the system as a service which has to be in the working mode and connected to the Internet, so the Logs are thrown and saved on some other web server termed as the evidence publisher not in the control of the CSP. So for the service to be always upright, we need to trust the CSP. The Logs L are captured for all the clients C who use the services S using the virtual

Table 1 Description of entities

#Sr-No	Set of Entities	Description
1	$C = \{c_1, c_2, c_3 \dots c_m\}$	m no of Clients
2	$S = \{s_1, s_2, s_3 \dots s_n\}$	n no of Services
3	$V = \{v_1, v_2, v_3 \dots v_k\}$	k no of Virtual Machines
4	$L = \{l_1, l_2, l_3 \dots l_p\}$	p no of Logs
5	$E = \{e_1, e_2, e_3 \dots e_q\}$	q no of Evidences

machines V . Such Logs termed as backup will serve as evidence of past logs termed as E in the threat model. Once the Logs enter the evidence publisher module, nobody can make changes to the published evidences termed as EPL.

6 Prototype Implementation and Results Observed

Our prototype environment configuration is illustrated in Fig. 3. The base operating system used for the implementation of Ubuntu 14.4. The OpenStack Icehouse is used to set up the Cloud and create the virtual machines represented as V in the threat model. Wireshark is used to capture the Logs. A few virtual machines with IP address of 10.0.47.1 and more are created as shown in Fig. 3 to capture the Logs represented as the L in the threat model. A client represented as C in the threat model uses these virtual machines as one of the services represented as S in the threat model provided by the CSP. The Logs generated by the system are encrypted using the elliptic curve cryptography (ECC) [25]. ECC is used to create faster, smaller, and more efficient cryptographic keys. It is an asymmetric key generation algorithm (one key for encrypting and the other for decrypting). A 256-bit ECC key is equivalent to RSA 3072 bit keys (which is 50% longer than 2048 bit keys used today). In terms of more traditional web server use cases, the smaller key sizes can offer speedier SSL handshakes (which can translate to faster page load times) and stronger security. As it generates asymmetric keys, one of the key can be used by the CSP to encrypt the logs and the other key can be used by the forensic investigator for decrypting when needed to verify the evidences. A hash of the Logs is computed using the cryptographic SHA512 hash function to form an integrity log chain also called as hash chaining. The computed hash value can be used to verify the integrity of copies of the original Logs without providing any original data. Hashing is one way not like encryption process in which decryption may be required. The Logs are signed by the CSP at the end of the day at a specified time. If not signed the CSP can be fined by the Law if required to make the CSP responsible.

The Logs captured would have the elements source IP (SrcIP), Destination IP (DstIP), time stamp (Ts) the time at which these web sites were accessed, Source Port (ScrPort), Destination Port (DstPost), Userid can be got from the Nova Database of the OpenStack. The Log details have been shown in Fig. 4. Few elements of the Log are encrypted using the ECC algorithm as discussed priorly termed as encrypted log entry (ELE). A hash chain (HC) is formed using the hashing algorithm which takes the ELE and HC of the previous record. At the initial step when the first log is taken, the HC value can be set to some assumed negative value as there will be no preceding log for the first log. A hash chain is formed for all the accesses made by the different clients while using the services as shown in Fig. 4. These logs are stored on the Insistent Database and also on the web server under control of some other CSP. Such multiple copies can be stored with different CSPs if required. The encrypted log plus the hash code of the previous Log encrypted gives a better security level for the Logs. As encryption is applied to the Log elements plus the hash code,

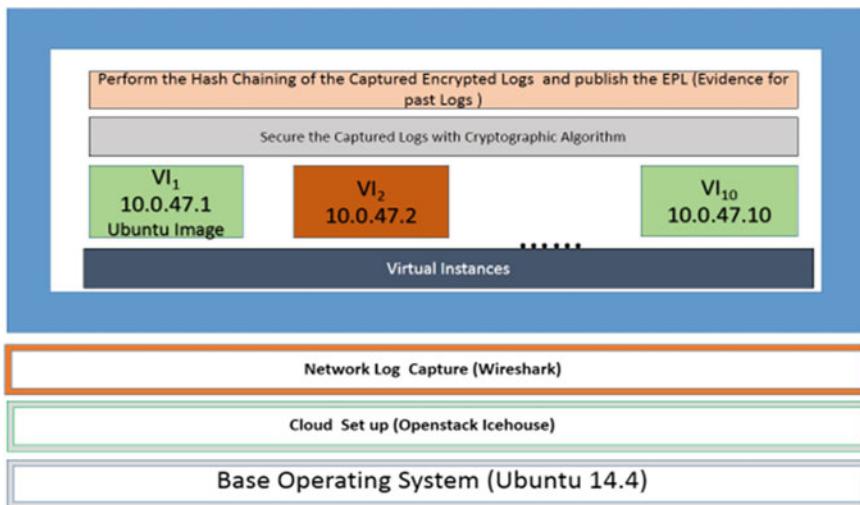


Fig. 3 Prototype environment configuration

confidentiality and integrity is also provided. As in the two keys generated, one of them is used by the CSP to sign for the encrypted elements and the other is shared with the forensic investigator and only they can view the data and no one else so the privacy is maintained and the authenticity is maintained. Multiple copies are kept at various web servers and the other CSPs if required, so there are many copies available. The forensic investigator when allocated the duty for analysis can check the hash value of the Log in case a crime is detected and the Logs on the server say the CSP's infrastructure was used to pursue the crime. One of the key is shared by the CSP with the forensic investigator. If the hash code of the Forensic Log in the backup copy is same with the one the CSP has then it concludes that it's not compromised till date and the logs are said to be authentic and the criminal can be sentenced to penalty as per the Law. CSP also will be alert in such cases and as the backs are available he cannot deny to give any data in raw form. Thus, if a crime is committed by using the hosted services S by the Client C and then closes the virtual machines V after committing the crime, the CSP has the data of the past logs EPL even if the virtual machines are shut off to hide the crime.

Python language has been used to code with MySQL in the backend for the data storage. The backup configuration system also has to be same as provided by us for the designed system to work properly. Database is used in the back end and to search as well. Searching such huge database can be quite an effort and will cost in time when partial data is known by the investigator but our main objective is to secure the Logs and provide evidence for past Logs. Different search functions are written which can be useful for the investigator to search for criminals Logs. The different observations are made as shown in Table 2, and the graph for the evidence for past

- LE = < SrcIP; DstIP; Ts; SrcPort; Dst Port; UserId >; ** Log Entry
- ELE= {EPKA(DstIP; DstPort; UserId,Ts); SrcIP; } ** Encrypted using ECC Log entry
- HC ={ H(ELE;HC_{Preceding}) } ** Hash Chain
- IDLE =< ELE;HC > ** Insistent Database Log Entry



Fig. 4 Encryption and hash chaining process for the logs captured

Table 2 Observed parameters

#Sr-No	Description	Time noted in milliseconds (ms)
1	Key pair generation time	17.659
2	Encryption time for a single Log	3.4029
3	Hashing time	0.0131
4	Encryption and Hashing time	3.4160
5	Database search time for an existing log	4.380

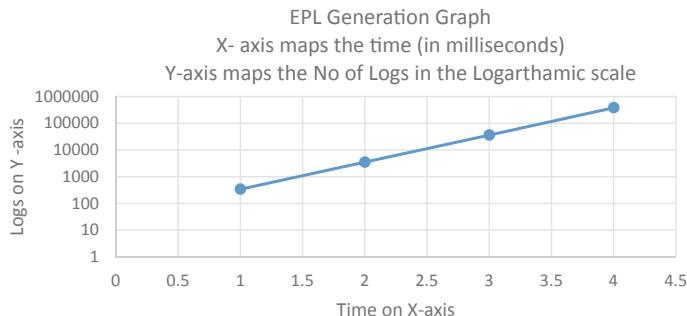


Fig. 5 EPL generation graph (encryption and hash chaining process) for the logs captured

logs is shown in Fig. 5. The graph observed is bound to be linearly increasing as the no of Logs increases the time also increases.

7 Conclusion and Future Work

A secured forensic Cloud framework is a need in today's world which aids in gathering expedient additional "evidences" against certain crimes which could be useful in framing the criminals in the court of Law in many countries like India. Security and forensics are compulsorily needed together as the better understanding and usage of security is a precondition for better indulgence and application of forensics. The system developed by us is a secure system with the different security parameters supported and justified. A CSP also will be alert with the compliance of standards and as the backups are available the CSP cannot deny to give any data in the raw form. Thus, if a crime is committed by using the virtual machines on the hosted services braced with our secure system, by the clients in India or any other country and then shuts them off after committing the crime, the CSP has the data of the past logs and even if the virtual machines are shut off to hide the crime the evidence lies with our system. If the virtual machines are closed without taking the Logs, and if the same virtual machines are allotted to some other users leading to loss of old data and thus losing important logs which would have the possible evidences. The backup Logs can be stored at some other CSPs or some other web servers and also can be used as evidence publisher. The result graph observed is bound to be linear as the time increases with the increase in the number of Logs which are going to be in millions on the server. Database search in million records can be very costly but nothing can be compromised for security. Future work can be focused by using different data structures to optimize the QoS parameters.

References

1. Ray, I., Belyaev, K., Strizhov, M., Mulamba, D., Rajaram, M.: Secure logging as a service—delegating log management to the cloud. *IEEE Syst. J.* **7**(2), 323–334 (2013)
2. Alqahtany, S.: A Forensically-Enabled IAAS Cloud Computing Architecture [online] <https://pearl.plymouth.ac.uk/bitstream/handle/10026.1/9508/2017alqahtany10426405phd.pdf?sequence=1&isAllowed=y>. Accessed last on 4 July 2018
3. Marturana, F., Tacconi, S., Italiano, G.F.: A forensic-as-a-service delivery platform for law enforcement agencies in cybercrime and cloud forensics. In: Applications for Investigation Processes. pp. 313–330, IGI Global, USA (2013). <https://doi.org/10.4018/978-1-4666-2662-1.ch013>
4. Spyridopoulos, T., Katos, V.: Data recovery strategies for cloud environments cybercrime and cloud forensics In: Applications for Investigation Processes. pp. 251–265, IGI Global, USA (2013). <https://doi.org/10.4018/978-1-4666-2662-1.ch010>
5. Pandi, G.S., Wandra, K.H.: Cloud forensic frameworks, challenges, state of art and future directions. *J. Emerg. Technol. Innovative Res.* **5**(5), 712–721 (2018)
6. Simou, S., Kalloniatis, C., Kavakli, E., Gritzalis, S.: Cloud forensics: Identifying the major issues and challenges. In: Advanced Information Systems Engineering, pp. 271–284. Springer, Berlin (2014)
7. Barrett, D.: Security architecture and forensic awareness in virtualized environments. In: Cyber-crime and Cloud Forensics: Applications for Investigation Processes. pp. 129–155, IGI Global, USA (2013). <https://doi.org/10.4018/978-1-4666-2662-1.ch006>

8. Patrascu, A., Patriciu, V.: Logging system for cloud computing forensic environments. *J. Control Eng. Appl. Inf.* **16**(1), 80–88 (2014)
9. Aydin, M., Jacob, J.: A comparison of major issues for the development of forensics in cloud computing. In: Proceedings of 8th International Conference for Internet Technology and Secured Transactions (ICITST-2013) (2014). <https://doi.org/10.1109/ICITST.2013.6750166>
10. Thorpe, S., Grandison, T., Campbell, A., Williams, J., Burrell, K., Ray, I.: Towards a forensic-based service oriented architecture framework for auditing of cloud logs. In: Proceeding of the IEEE 9th World Congress on Services, IEEE 2013 Services Workshop on Security and Privacy Engineering (SPE2013), pp. 75–83. Santa Clara, CA, USA (2013)
11. Trenwith, P.M., Venter, H.S.: Digital forensic readiness in the cloud. In: Information Security for South Africa. pp. 1–5, IEEE (2013)
12. Zawoad, S., Mernik, M., Hasan, R.: Towards building a forensics aware language for secure logging. *Comput. Sci. Inf. Syst.* **11**(4), 1291–1314 (2014). <https://doi.org/10.2298/csis131201051z>
13. Zawoad, S., Dutta, A.K., Hasan, R.: SecLaaS: Secure logging-as-a-service for cloud forensics. In: Proceedings of the 8th ACM Symposium on Information, Computer and Communications Security (ASIACCS). pp. 219–230, ACM, New York (2013)
14. <https://www.globalsign.com/en/blog/elliptic-curve-cryptography/> Accessed on 4 July 2018.
<http://cse.iitkgp.ac.in/~debdeep/pres/TI/ecc.pdf> Accessed last on 18 Apr 2018
15. Zawoad, S., Dutta, A.K., Hasan, R.: Towards building enables cloud through secure logging-as-a-service. *IEEE Trans. Dependable Secure Comput.* **13**(2), 148–162 (2016)
16. Dykstra, J., Sherman, A.T.A.: Understanding issues in cloud forensics: Two hypothetical case studies. In: Proceedings of the 2011 ADFSL Conference on Digital Forensics Security and Law. pp. 1–10 (2011)
17. Dykstra, J., Sherman, A.: Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform. *Digit. Invest.* **10**, S87–S95 (2013)
18. Sang, T.: A log-based approach to make digital forensics easier on cloud computing. In: Proceeding of 3rd International Conference on Intelligent System Design and Engineering Applications. pp. 91–94 (2013). <https://doi.org/10.1109/ISDEA.2012.29>
19. Marangos, N., Rizomiliotis, P., Mitrou, L.: Digital forensics in the cloud computing era. In: GC'12 Workshop: 1st International Workshop on Management and Security Technologies for Cloud Computing, pp. 775–780, IEEE (2012). 978-1-4673-4941-3/12
20. Simon, B.: ISO standards and computer forensics. *Forensic Focus* 2018, [online] Available: <http://www.forensicfocus.com/index.php?name=Content&pid=343>. Accessed last on 4 July 2018
21. Patrascu, A., Patriciu, V.: Beyond digital forensics. A cloud computing perspective over incident response and reporting In: Applied Computational Intelligence and Informatics (SACI), pp. 455–460 (2013)
22. Mell, P., Grance, T.: Nist cloud computing forensic science challenges. In: Draft NISTIR 8006, June (2014)
23. Khan, S., Gani, A., Wahab, A.W.A., Bagiwa, M.A., Shiraz, M., Khan, S.U., Buyya, R., Zomaya, A.Y.: Cloud log forensics: Foundations, state of the art, and future directions. *ACM Comput. Surv.* **49**(1), 7:1–7:42, Article 7, 42 (2016). DOI: <https://dl.acm.org/citation.cfm?doid=2911992.2906149>
24. Alqahtany, S., Clarke, N.L., Furnell, S., Reich, C.: Cloud forensics: a review of challenges solutions and open problems. In: Proceedings of Cloud Computing (ICCC) International Conference (2015)
25. Zawoad, S., Hasan, R., Skjellum, A.: Towards achieving reliable digital forensics in IaaS and Stass Clouds using the open cloud forensics model. *Serv. Trans. Cloud Comput.* **4**(3) July-September (2016). (ISSN 2326-7550)

26. Jahdali, H.A., Albatli, A., Garraghan, P., Townend, P., Lau, L., Xu, J.: Multi-tenancy in cloud computing. In: Proceeding of the IEEE 8th International Symposium on Service Oriented System Engineering, pp. 344–351. Oxford, United Kingdom (2014)
27. Ruan, K., Carthy, J., Kechadi, T., Crosbie, M.: Cloud forensics: An overview In: Proceedings of the 7th IFIP International Conference on Digital Forensics (2011)

Document Classification Using Wikidata Properties



Sheeban Wasi, Madhurendra Sachan and Manuj Darbari

Abstract The recent advancements in computer science and technology additionally lead to the generation of a prodigious amount of documents in digital form which are perpetually incrementing. They are widespread in structured as well as semi-structured format. The documents are interrelated and convey the same message but broadcasted from different sources. In this paper, we have endeavoured to enhance the document classification by using classification algorithms integrating with external knowledge base from Wikidata and classify the document of same type in appropriate categories using Wikidata properties as features.

Keywords Wikidata · tf-idf · Learning algorithms · Feature selection · VSM · Tokenisation · Classification

1 Introduction

Classification is a necessity of any system which processes documents; a fundamental step is to build knowledge base for a model by training certain algorithms, but as we move towards social media or news media data, we cannot limit the system to certain categories. The news data is growing at a rapid rate, and getting piled up in digital repositories, it is difficult to access the information efficaciously, so one way to solve this quandary is automatic categorisation [3].

Classification is an imperative technique to label a document based on the relevant content of the document. For example, Apple iPod can be classified in ‘MP3

S. Wasi · M. Sachan (✉)
School of Engineering, BBD University, Lucknow, India
e-mail: madhurendra@tikaj.com

S. Wasi
e-mail: sheeban@tikaj.com

M. Darbari
Department of IT, BBDNITM, Lucknow, India
e-mail: manuj_darbari@acm.org

player category' and 'Apple Products'. In growing social data, a classifier needs to constantly use a knowledgebase system which is improved daily. For example, if someone is talking about 'xyz' which is newly launched music, it should know about 'xyz' as music. Some of the classifiers which we used are multinomial Naive Bayes classifier, support vector machines, logistic regression, ridge classifier, etc. But, before using these algorithms, pre-processing is required which also helps in improving the classification significantly [4, 9]. We also used vector space model which is an important part of this research.

Vector space model (VSM) is an approach which is widely used for representation of the document as a point in space. The value is decided by the frequency of the terms in a particular document. These words are then used as features which are called bag-of-words. But the challenge here is dimension because the documents contain high-dimensional texts which extend the processing time. So, feature extraction needs to be done such that only the important terms get trained reducing the high dimensionality for increasing the efficacy of the classification process. Classification results can be affected by the redundant data termed as noisy data, so pre-processing is required in order to remove it—Stop word removal and stemming are performed. To find the weight of the word in the document and convert the document in a structured format, a scheme called *tf-idf* (term frequency-inverse document frequency) is used; the *tf-idf* value is directly proportional to the weight of the terms in the collection of documents, but is offset by the frequency of the words in the corpus, which helps to determine words that are more useful or common as compared to others. This can amend categorisation as it provides semantic background knowledge to document. Semantic background knowledge can additionally be extracted through WordNet, Wikidata etc.

In this paper, we propose a way to amend semantic knowledge by utilising Wikidata. We endeavour different coalescences to enrich documents, utilising Wikidata properties. Our experiment shows that utilising only Wikidata as our knowledge base (without *tf-idf*) gave remarkable results.

2 Previous Work

There are numerous researches conducted in the field of digital document classification, and a lot of algorithms have been implemented in order to create a classifier. In the previous approach for categorisation, the features from within the textual document are extracted and used for performing classification [1, 8]. For effectively determining the class relationship between the article and the category, words, phrases, sequences and concepts perform a crucial role. But this is limited and cannot be used as a core idea for classification algorithm. So in order to overcome this limitation, numerous researches proposed an idea to introduce an external knowledge base to enrich the document representation [4, 5, 10]. Researchers used Open Directory Project (OPD) [6] and WordNet [1] to enhance the knowledge and enrich the document representation. OPD fails to give acceptable results due to unorganised

hierarchy and its extremely unbalanced nature. WordNet is the thesaurus of English language, and in various models, the general concepts are extracted from WordNet.

3 Proposed Method

Training a classifier and using Wikidata as knowledge base, primarily we need to pre-process the data. We used standard data sets used by research community so that we can compare the quality of our work. The data set we used is BBC data set to further check the performance of our models.

3.1 *Pre-processing*

The most imperative part is to pre-process the data so that further implementations can be done. The basic processing steps followed are listed which helps in making the process efficient by reducing the process time.

Tokenisation

Text mining research usually involves words or sentences which splits word by word to process it any further. So in this, all the words will be split and the punctuations will also be discarded since it cannot represent any category.

Example:

Input: Donald Trump Is Elected President

Output: [Donald] [Trump] [Is] [Elected] [President].

Stop Word Removal

Stop word removal has been used by many researchers in the field of text mining for information retrieval. This algorithm removes the words which are not necessary to define the category of the article, or we can say unimportant words like conjunctions and prepositions that only increase the process time and making the calculations complicated. This removal process uses the stop word dictionary.

Example: ‘a’, ‘an’, ‘the’, ‘I’, ‘an’, ‘will’ or any other prepositions and conjunctions.

Input: Donald Trump is the elected President,

Output: [‘Donald’, ‘Trump’, ‘Elected’, ‘President’].

N-gram Creation

In order to handle multiword entities and reduce false lookups, we used n-gram where $n = 1, 2, 3, 4$. We excluded subsets if certain set was discovered in knowledge base.

Input: [‘Donald’, ‘Trump’, ‘Elected’, ‘President’, ‘Stunning’, ‘Repudiation’, ‘Establishment’]

Output: (Partial):

- donald-trump-elected-president, donald-trump-elected, donald-trump, Donald, trump, trump-elected, trump, elected.

Above processing would exclude ‘donald’ and ‘trump’ if ‘donald-trump’ is found in knowledge base. Complexity of search method is

$$O(n) = n!$$

where n is number of words; the search space could be further reduced by not looking up word pairs which had stop words in between.

Term Frequency–Inverse Document Frequency

It is a statistical measure to calculate how important a word is in a corpus. Term frequency is the number of times a word can be found in an article, and idf is the computation from log of the inverse probability of word being found in any article.

$$idf_i = \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (1)$$

$|D|$: cardinality of D , or the total number of documents in the corpus $|\{j : t_i \in d_j\}|$: number of documents where the term t_i appears (viz. the document frequency) (that is $n_{i,j} \neq 0$). If the term is not in the corpus, it will lead it to division-by-zero. Therefore it is common to use

$$1 + |\{j : t_i \in d_j\}| \quad (2)$$

1. Donald Trump is the president of USA.
2. Indian president and Israel president were a part of the international summit.
3. Elections are five years ahead from now.

Let us assume we want to calculate the weight of the term ‘president’ in the document in #1 and #2. As we can see in the list that ‘president’ comes once in #1 and twice in #2 and the term is written in two documents, the calculation can be seen as follows:

- $Tf-idf(\text{learning}, 1) = 1 * \log(3/2) = 0.23$
- $Tf-idf(\text{learning}, 2) = 2 * \log(3/2) = 0.46$.

POS Tagging

Part-of-speech tagging is a method to annotate words in a sentence as noun, verb, adjective, etc. For this work, we use MaxentTagger which is provided as part of Stanford Core NLP, with default model, i.e. left3words-wsj-0-18.tagger.

Example:

Input: Donald Trump Is Elected President in Stunning Repudiation of the Establishment

Output: [('Donald', 'NNP'), ('Trump', 'NNP'), ('Is', 'VBZ'), ('Elected', 'VBN'), ('President', 'NNP'), ('in', 'IN'), ('Stunning', 'NNP'), ('Repudiation', 'NNP'), ('of', 'IN'), ('the', 'DT'), ('Establishment', 'NN')].

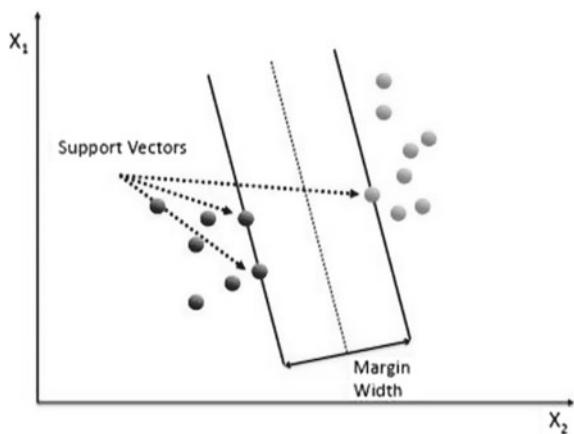
Here, NNP, VBZ, VBN, IN and DT are standard part-of-speech tags.

3.2 Classification Algorithm

Classification algorithm is a supervised learning method in which the model learns from the data input given and uses the learning to classify new observation. We used classification algorithms to improve the model. SVM gave much better results as compared to other classification algorithms. SVM algorithm was proposed by Vapnik [7] in order to get a solution for two-class problem.

Basically, SVM finds a separation between hyperplanes which is defined by classes of data. It is trained using the pre-classified documents and gives good results as compared to Naive Bayes and other alternatives when the entire vocabulary was used as the feature set because it can operate in fairly large data sets as it measures the margin of separation rather than matches on features. Different semantic functions can be used as kernels K: radial basis and polynomial basis kernels which affect the accuracy of the algorithm (Fig. 1).

Fig. 1 SVM hyperplane



3.3 Wikidata Integration

Wikidata [11] is an ontological representation of Wikipedia documents, which are generated dynamically but can also be manually updated. The data consists of document—where each item represents an object or a property.

In a classification problem, some contextual information can help in classification by adding the property specific to that article/document [2, 12, 13].

Example:

- Business news can contain more items which are organisation, company, product, CEO, founders, investors, etc.
- An entertainment article would contain entities related to movies, actors, etc.
- A technology-related cluster would talk more about items related to applications, mobile phones, development, etc.

Some examples to items which can be found during analysis:

<https://www.wikidata.org/wiki/Q5284>

<https://www.wikidata.org/wiki/Q189490>

<https://www.wikidata.org/wiki/Q95>

<https://www.wikidata.org/wiki/Q22686>.

As initial processes we imported Wikidata to MongoDB which is document-based—schema-less database. It contained 45386255 documents.

4 Methodology

The pipeline consisted of methods to reduce the search space. The pipeline had the following stages for:

1. Load all ‘en’ valued labels and aliases from Wikidata into memory. Normalise each value to a base form. e.g.: Donald Trump to donald-trump
Store all references found in a hashmap.
2. POS tag each word in the input text, and extract all proper nouns in order.
e.g.: ['Donald', 'Trump', 'President', 'Stunning', 'Repudiation'].
3. Create n -gram for extracted list of proper nouns, using sliding window approach.
 $N = 4, 3, 2, 1$
e.g.: ['donald-trump-president-stunning', 'donald-trump-president', 'donald-trump', 'donald', 'trump', 'trump-president', 'trump', 'trump-president-stunning', 'trump-president', 'trump', 'president' ...] and so on.
4. Normalise the pairs to a base form and look for their presence in loaded hashmap.
Words found in Wikidata:
donald-trump(Q5295230, Q23001025, Q22686, Q5295229, Q27947481),
president ('Q1255921' 'Q7241206', ...)
stunning (Q7628762)

Table 1 Results of the proposed method

Category	Precision	Recall	F1 scores	Support
Business	0.99	0.96	0.97	122
Entertainment	0.96	0.97	0.96	69
Politics	0.95	0.97	0.96	73
Sports	1.00	0.99	1.00	104
Technology	0.96	0.99	0.97	77
Avg/Total	0.98	0.98	0.98	445

Accuracy: 96.8492

Table 2 Results of SVM with tf-idf

Category	Precision	Recall	F1 scores	Support
Business	0.94	0.94	0.94	95
Entertainment	0.97	0.95	0.96	75
Politics	0.92	0.91	0.91	88
Sports	0.95	1.00	0.98	106
Technology	0.96	0.94	0.95	81
Avg/Total	0.97	0.97	0.97	445

Accuracy: 95.9551

repudiation ('Q19358049', 'Q21071664').

5. Use frequency of occurrence of Wikidata properties in found words as feature in classifier.
In preceding example following:
We count occurrences of Q5295230, Q23001025, Q22686, Q27947481, Q1255921, Q7241206 in given text input.
6. Remove noisy features with negligible variance.
7. Since there are over 4000 + Wikidata properties, we apply latent Dirichlet allocation (LDA) for feature reduction with $N = 6$ (N —number of topics).
8. Train multiclass SVC classifier.

5 Results

We used BBC data set for training and testing. Since BBC data set is very huge, we created a sub-data set of 3000 samples, where we used 80% for training and 20% for testing.

We used Jaccard similarity score for generating accuracy score, which is basically size of intersection of actual classification and predicted results for each index divided by size of union of actual classification and predicted (Tables 1 and 2).

Above results clearly reflect that the knowledge base affects the accuracy of prediction, depending on category (Figs. 2 and 3).

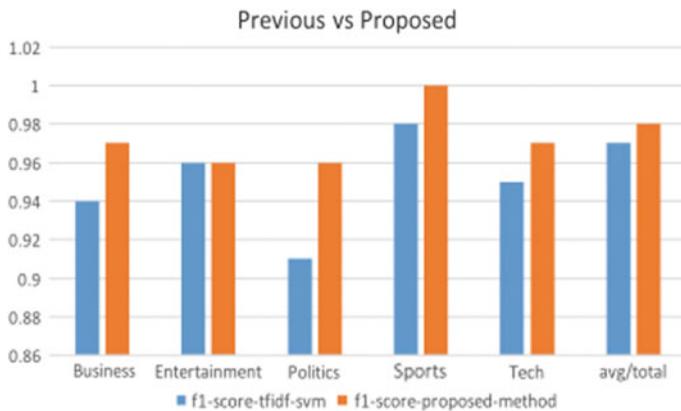


Fig. 2 F1 score comparison

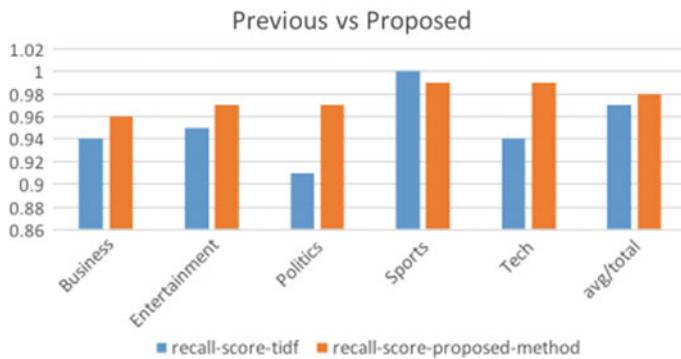


Fig. 3 Recall score comparison

6 Conclusion and Future Work

The proposed method has increased the accuracy of data classification to 96.84% (about 2% above the conventional method). At the same time, the method reduces the dependency on training data sets and uses an external knowledge base. The method by utilising a dynamic data set enhances the classification of real-world news and articles.

This method can in future be evolved to autonomously create clusters when being trained on generic parameters. Future researchers can also use this method to automatically classify and create clusters of data from varied sources in real world.

References

1. Elberrichi, Z., et al.: Using WordNet for text categorization. *Int. Arab J. Inform. Technol. (IAJIT)* **5** (2008)
2. Hassan, S., et al.: Comparing SVM and naïve Bayes classifiers for text categorization with Wikitology as knowledge enrichment. In: 2011 IEEE 14th International Multitopic Conference (2011)
3. Ittoo, A., et al.: Text analytics in industry: challenges, desiderata and trends. *Comput. Ind.* **78**, 96–107 (2016)
4. Khan, M.T., et al.: Sentiment analysis and the complex natural language. *Complex Adapt. Syst. Model.* **4**, 15 (2016)
5. Muliono, Y., Tanzil, F.: A comparison of text classification methods k-NN, Naïve Bayes, and support vector machine for news classification. *JPIT* **3**, 157–160 (2018)
6. Osiński, S., Weiss, D.: Conceptual clustering using lingo algorithm: evaluation on open directory project data. In: Intelligent Information Processing and Web Mining, pp. 369–377 (2004)
7. Sebastiani, F.: Machine learning in automated text categorization. *ACM Comput. Surv.* **34**, 1–47 (2002)
8. Sun, A., et al.: Web classification using support vector machine. In: Proceedings of the Fourth International Workshop on Web Information and Data Management (WIDM '02) (2002)
9. Talib, R., et al.: Text mining: techniques, applications and issues. *IJACSA* **7** (2016). <https://doi.org/10.14569/ijacsa.2016.071153>
10. Tripathy, A., et al.: Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* **57**, 117–126 (2016)
11. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. *Commun. ACM* **57**, 78–85 (2014)
12. Wang, P., et al.: Improving text classification by using Encyclopedia kKnowledge. In: Seventh IEEE International Conference on Data Mining (ICDM) (2007)
13. Zhang, Z., et al.: Improving semi-supervised text classification by using Wikipedia knowledge. In: Lecture Notes in Computer Science, pp. 25–36 (2013)

Energy-Efficient Fuzzy-Logic-Based Data Aggregation in Wireless Sensor Networks



Sukhchandan Randhawa and Sushma Jain

Abstract Wireless sensor networks have limited processing capability and limited battery power. Due to large collection of input data, it is difficult to manage the data along with different domains. Hence, energy-efficient data aggregation technique is required for efficient data collection. Data aggregation is the method in which data coming from different sensors is combined and provides useful aggregated information. Keeping in view the above issue, a novel energy-efficient fuzzy-logic-based data aggregation technique is proposed. The proposed technique collects, analyzes, classifies, and aggregates the data of different domains automatically which is reported by various sensors. Further, fuzzy logic technique is applied as it has capability to deal with dynamic situations and to model the conditions which are inherently imprecisely defined. The proposed data aggregation technique aggregates the incoming data in an effective manner by reducing energy consumption based on different fuzzy rules designed in knowledge base, which further improves network lifetime. The performance of the proposed technique has been evaluated and compared with the existing technique, i.e., energy-efficient scheduling strategy (EESS) in terms of energy consumption, data aggregation rate, data persistence, and network lifetime.

Keywords Wireless sensor networks · Networking · Fuzzy logic · Data aggregation

1 Introduction

In wireless sensor networks (WSNs) [1, 2], the requirement of energy for computation purpose when compared to that of required for communication is insignificant. “The process of collecting and combining the useful information is called data aggrega-

S. Randhawa (✉) · S. Jain

Computer Science and Engineering Department, Thapar University, Patiala, India
e-mail: Sukhchandan@thapar.edu

S. Jain

e-mail: sjain@thapar.edu

tion” [3]. Using data aggregation, a large amount of energy can be saved. The gathered data should be aggregated efficiently so that it can be aggregated in future for different types of forecasting. But there is no effective data aggregation technique exists which can aggregate data with minimum energy consumption, maximum data aggregation rate, data persistence, and network lifetime. To solve this problem, energy-efficient data aggregation technique is therefore required which improves network lifetime.

The motivation of this paper is to design a fuzzy-logic-based data aggregation (FLDA) methodology. This methodology can be utilized for gathering, analyzing, classifying, and aggregating multidimensional data sent by deployed sensors. The primary idea behind the presented work is: (1) utilizing sensors for data collection; (2) cluster formation of data sensed; (3) effectively aggregating data, minimizing energy consumption, maximizing data aggregating rate, data persistence, and network lifetime. The structure of this paper is as follows: Section 2 focuses on contributions and related work. The proposed methodology is stated in Sect. 3. The setup for experimental and performance analysis is showcased in Sect. 4. The future directions and conclusion are discussed in Sect. 5.

2 Related Work

The literature reports different methods of aggregating data on which research is going on presently, namely cluster-based, energy-based, latency-based, network-lifetime-based, and QoS-based data aggregation. In cluster-based data aggregation, there is an elected cluster head (CH) within each cluster to collect the data of member nodes. Jun et al. [4] presented the method of distributed aggregation where application of Slepian-Wolf code is made for selecting a potential disjoint cluster. A clustering protocol is proposed with a negligible complexity joint coding method for reducing the redundancy of data due to intercluster spatial correlation. The Hybrid-Based_Data_Aggregation (HBDA) [4] uses numerous clustering methodologies simultaneously, and the most adequate is selected based on the status of the network. In initial stages, the generation of tree topology takes place among the sink and the member nodes. In successive phases, cluster head selection takes place using static selection method. Two-tier cluster-based data aggregation (TTCDA) [6] method makes use of functions based on the additive and divisible aggregation of generated packets by nodes utilizing temporal–spatial correlation. Aggregation is performed at two levels: An intracluster data aggregation is performed by local aggregators (LA), which transmit the data aggregated to gateway/aggregator nodes (A/G).

Energy is one of the important QoS parameters for effective aggregation of data. The cooperative communication-based (CCB) technique presented by Hongli et al. [7] comes out to be a problem which is NP hard in nature. In order to solve this, an approach based on heuristics known as Monte Carlo tree (MCT) is utilized. For communicating cooperatively, a single transceiver node is deployed and many other nodes can transmit their data to it. For alleviating the issue of conflicts between

members while transmitting data, Yingshu et al. [8] proposed latency-based and distributed aggregation method. It leads to generation of cluster-based formation of tree and their cluster heads, respectively, for improving data transmission among clusters. To obtain optimal aggregation trees for solving problems, a compressed data aggregation considering the principle of joint routing and compressed aggregation was proposed [9]. This is based on greedy heuristic and mixed integer programming formulation (GHMIPFC). Hongjuan et al. [10] proposed a method of data aggregation based on accuracy of data without revealing private node readings and making a reduction in overhead. Hongxing et al. [11] proposed distributed data aggregation algorithm (DDAA) utilizing the model of physical interference and $O(K)$ slots of times for completing the aggregation process. Here, K represents the logarithmic value of ratio of shortest to longest length of link in the network. A network partitioning is done into cells utilizing the principle of nearest neighbor. Konstantinos et al. [12] designed a method to increase energy efficiency. Here, the capturing of data is done from all sensor nodes in a systematic manner and is propagated to the sink. Xueyan et al. [13] proposed a method based on precision constraint known as CBDA. For differentiating the precisions, the quotient of precision constraint and error bounds are taken in dynamic fashion for improving lifetime of the network.

In *data aggregation on the basis of QoS*, various QoS parameters such as energy, delay, accuracy, fault tolerance, and network lifetime are considered based on the underlying application. Chen et al. [14] presented a data aggregation method on the basis of adaptive fault tolerance (AFT) in which a mathematical model is proposed for satisfying requirements of QoS. Sudip et al. [15] proposed a data aggregation technique in which data is transmitted via shortest path between source and destination. Information is stored by each node about its neighboring node only to save memory. The aggregation of data is executed by following node of source node. Energy-efficient scheduling strategy (EESS) [16] utilizes TDMA technique for the tasks of a subset of nodes in varied groups in consecutive slots of time. EESS is ineffective in transferring complete data in time slots assigned because of data variation with respect to time. Hence, the sensor needs to be in awake and listening mode until all its remaining packets are sent, leading to unwanted energy wastage.

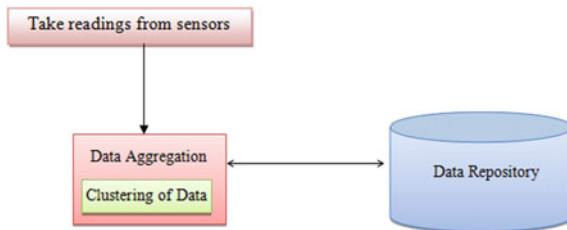
The novel FLDA technique collects the data, analyzes, classifies, and aggregates the data with minimum energy consumption and maximum data aggregation rate, data persistence, and network lifetime, which are not considered in other existing works. The comparison of FLDA has been done with some of the already presented methods given in Table 1.

Table 1 Comparison of FLDA with existing data aggregation techniques

Type of DA	Subtype	Process of aggregation of data	Parameter(s)
Cluster	CBADA [5]	Spatial aggregation and temporal aggregation	Aggregation ratio
	HBDA [7]	Clustering-based aggregation	Fault tolerance
Latency	DDAA [15]	Nearest neighbor criterion	Latency and lifetime of the network
Energy	CCB [10]	Monte Carlo tree	Energy
	GHMIPFC [12]	Joint routing and compressed aggregation	Energy
NL	CBDA [18]	Precision constraint partitioning	Energy efficiency and network life span
QoS	AFT [14]	Clustering-based query processing	Energy and lifetime of the network
FLDA (proposed)		QoS-aware fuzzy logic-based data aggregation	Energy consumption, data aggregation rate, data persistence, and network lifetime

3 Energy-Efficient Fuzzy-Logic-Based Data Aggregation

Data aggregation on the basis of fuzzy logic (FLDA) has been proposed whose primary focus is to provide effective aggregation of data for improving data aggregation rate, data persistence, and lifetime of the network and to reduce energy consumption. Figure 1 shows the process of data aggregation. Data aggregation model works in following steps: (i) sensors are used to capture the data, (ii) clustering of gathered data is done, (iii) data aggregation in an energy-efficient manner is performed and data is stored in repository for future use.

**Fig. 1** Process of data aggregation

3.1 Clustering of Data

The utilization of K -means-based clustering is done for clustering the uncategorized data gathered by member nodes. This methodology takes population data (Data Set (D)) as input and then clusters this data into varied clusters on the basis of the value of the centroid. In order to decide the required count of clusters, the distance among the centroids of the selected points is considered. These centroids should be at the farthest distance possible from each other.

The data point is assigned to one of the clusters on the basis of its minimum distance from the centroid of the cluster (C). Till the time when all the data points in Data Set (D) are assigned to some cluster, the process is executed iteratively. The final clusters of data (CD) are generated. WEKA Tool [17] is used to perform clustering of data.

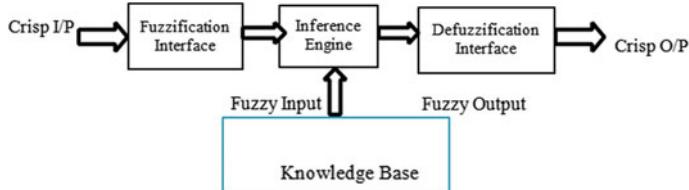
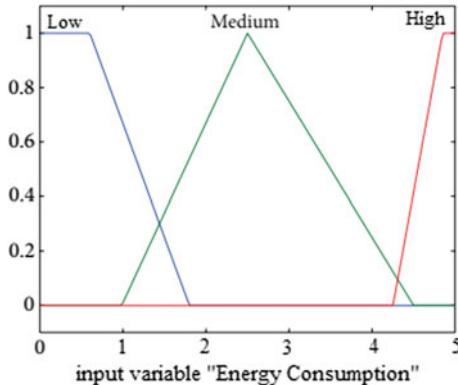
3.2 Fuzzy Logic in Data Aggregation

Figure 2 presents the main components of fuzzy system used in the proposed framework. The fuzzy system [18, 19] components are explained as follows:

- (a) ***Input–Output variables of Fuzzy System***: Data aggregation rate (DAR), energy consumption (E_k), and data persistence (DP) act as fuzzy inputs, and network lifetime (NL) comes out to be fuzzy output. There are three categories of input parameters: *Low*, *Medium* and *High*. The classification of all input and output variables is done into varied groups. Each group depicts set of given input or output fuzzy in nature. Table 2 shows fuzzy systems' input and output variables.
- (b) ***Fuzzy rule base***: For each of the three inputs, four membership functions are utilized. There are 81 rules for this system (3 inputs for each of the four variables (3^4)). The number of rules is user defined in inference process as given below.
If (E_k) is High and (DAR is Low) and (DP is Low) then (NL is Low).
- (c) ***Membership functions of Fuzzy System (Inference(s))***: For every input and output variables: Low (L), Medium (M) and High (H), membership functions are considered. Based on the conditions and requirements of every domain in which data is aggregated, the values of membership functions can be varied. The partitioning of these parameters should be done into suitable categories conceptually. The creation of member functions is done on the basis of fuzzy system's selected input and outputs, for better representation of relationship among them. The conceptual partitions are utilized for the creation of set of fuzzy rules as shown in Fig. 3. The membership degree of each fuzzy variable is represented by the y-axis. The quantized values represent the universe of discourse (x axis) for the input and output fuzzy variables.
- (d) ***Fuzzification***: For evaluating the rules and producing another output variable, fuzzy “AND” operator is utilized. It returns the least value among the entered values. The inference process applies the truth value of each rule to the output

Table 2 Fuzzy input and output variables

(E_k)	Data aggregation rate	Data persistence	Network lifetime
L	L	L	L
M	M	M	M
H	H	H	H

**Fig. 2** Components of fuzzy system**Fig. 3** Fuzzy subsets of fuzzy inputs and fuzzy output

value (R). For calculating the truth degree throughout the usage of fuzzy logic “AND” the inference rule is known as MINIMUM.

$$\text{AND } ((E_k), DAR, DP)) = \text{MINIMUM} (\text{truth}(E_k), \text{truth}(DAR), \text{truth}(DAR))$$

- (e) **Defuzzification:** The fuzzy output value is converted into crisp output value. In the proposed work, MAXIMUM technique is used for defuzzification. As per MAXIMUM technique, the variable value at which the fuzzy subset obtains its maximum truth value is selected as the crisp value for the output.

4 Experiment Setup and Results

To validate the proposed FLDA technique, EESS [16] is selected which also performs data aggregation. The comparison of their experimental results has been done for three distinct parameters, namely DAR, E_k , DP. The path with minimum (E_k) is decided by fuzzy member functions, and data is transmitted. The evaluation of FLDA is done through MATLAB and FuzzyJ Toolkit used for Java [20] to implement the necessary fuzzy logic functionality. Table 3 describes the simulation parameters used in this experiment work. That is, α denotes the energy dissipated in the transmission of message from source to destination.

Test Case 1: Energy Consumption (E_k): The value of (E_k) is calculated for both FLDA and EESS with distinct count of nodes (50–300). Energy is consumed by every sensor node for computing and communicating; therefore, as the node count increases so does the consumption of energy. E_k in FLDA is lesser as compared to EESS at different number of nodes as shown in Fig. 4. The minimum value of (E_k) is 0.5 J at 50 nodes.

Test Case 2: Data Aggregation Rate (DAR): DAR is known to be the ratio of the amount of successfully aggregated data to data sensed in total. As the amount of

Table 3 Simulation parameters

Parameter	Value
Network coverage/m ²	200 × 200 m ²
Number of sensors	50–300
Initial energy/J	0.5
Data packet size/Bit	500
Control packet size/Bit	12
A	2.5
Maximum transmission range/m	30

Fuzzy inputs:

Energy consumption (E_k): (from 0.5 to 5 J)

Data aggregation rate: (from 25 to 95%)

Data persistence: (from 0 to 1)

Fuzzy outputs:

Network lifetime: (from 10 to 100 s)

Fig. 4 (E_k) versus number of nodes

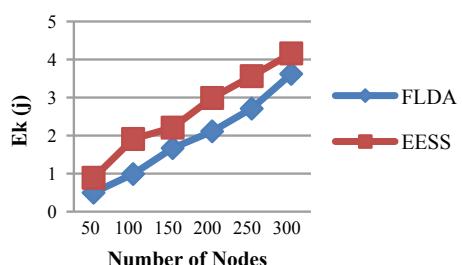


Fig. 5 DAR versus data amount

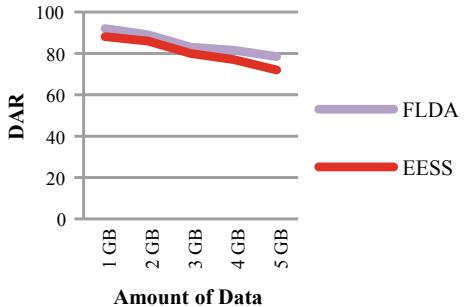
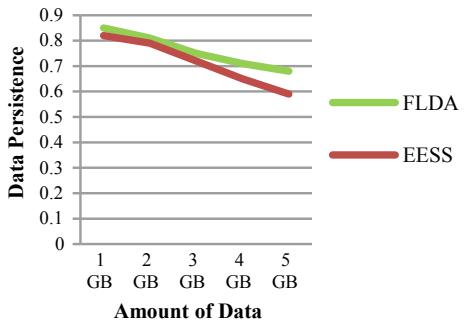


Fig. 6 DP versus data amount



data increases for aggregation, DAR shows decrement in value as shown in Fig. 5. FLDA shows better performance in measures of DAR than EESS. Data aggregation rate in FLDA is 8.47–13.11% more than EESS.

Test Case 3: Persistence of Data: Data persistence (DP) is known to be the ratio of amount of successfully backed up data on data repository to total amount of received data. As the amount of data increases for aggregation, DP shows decrement in value as shown in Fig. 6. The calculated value of DP for FLDA and EESS is determined, but FLDA has an upper hand over EESS in terms of performance. DP shows the highest value of .85 at 1 GB data in FLDA.

Test Case 4: Lifetime of the Network (NL): The value of NL is calculated for FLDA and EESS both with different number of nodes. The value of NL is decreasing, with the increase in the number of nodes. The value of NL in FLDA is greater in comparison with EESS for distinct node count as depicted in Fig. 7. The highest lifetime of the network is 96 s at 50 nodes.

Test Case 5: Convergence of FLDA: Figure 8 plots the convergence of total (E_k) computed by FLDA model over the number of iterations for different value of NL: 50, 100, 150, 200 s by aggregating data coming from different nodes. Initially, the data is randomly initialized. Therefore, the total initial energy consumption is very high at 0th iteration. As the model progresses, the convergence is drastic and achieves global minima very quickly. The number of iterations required for the convergence is seen to be 45–60, for simulated environment.

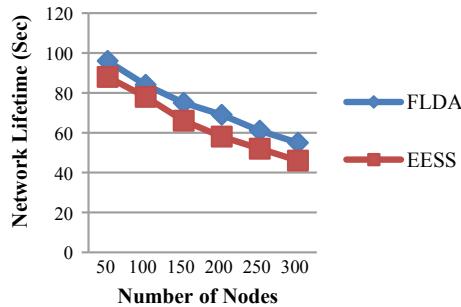


Fig. 7 NL versus number of nodes

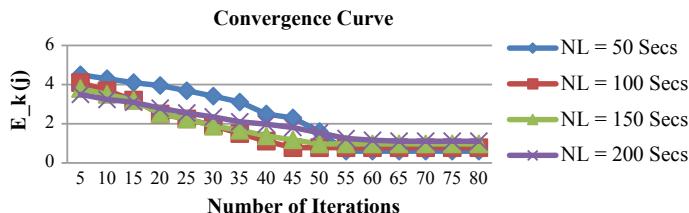


Fig. 8 Convergence of FLDA model

5 Conclusions and Future Scope

A fuzzy-logic-based data aggregation (FLDA) method for WSNs has been presented. The primary objective of the presented approach is reduction of the complexity of aggregation of gathered data. K -means-based clustering method is used for clustering the enormous uncategorized data reported by various sensors in continuous fashion. Further, fuzzy logic is applied to perform data aggregation based on rules. The result of simulations demonstrates that the proposed FLDA technique is effective in decreasing energy consumption and improving data aggregation rate, data persistence, and network lifetime as compared to the existing data aggregation technique, i.e., EESS as shown in test cases. In future, FLDA technique can be made autonomic to perform data aggregation automatically.

References

- Randhawa, S.: A Review of Power Aware Routing Protocols in Wireless Sensor Networks, pp. 22–23 (2012)
 Randhawa, S.: Research challenges in wireless sensor network: A STATE OF THE PLAY. In: National Conference on Convergence of Science, Engineering & Management in Education and Research, Moga, India, pp. 1–4 (2014)

- Randhawa, S., Jain, S.: Data aggregation in wireless sensor networks: previous research, current status and future directions. *Wirel. Pers. Commun.* **97**, 3355–3425 (2017)
- Zheng, J., Member, S., Wang, P., Member, S.: Distributed data aggregation using slepiant-wolf coding in cluster-basedwireless sensor networks. *IEEE Trans. Veh. Technol.* **59**, 2564–2574 (2010)
- Jung, W.-S., Lim, K.-W., Ko, Y.-B., Park, S.-J.: Efficient clustering-based data aggregation techniques for wireless sensor networks. *Wirel. Networks* **17**, 1387–1400 (2011)
- Mantri, D., Prasad, N.R., Prasad, R., Ohmori, S.: Two tier cluster based data aggregation (TTCDA) in wireless sensor network. In: IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp. 117–122 (2012)
- Xu, H., Huang, L., Zhang, Y., et al.: Energy-efficient cooperative data aggregation for wireless sensor networks. *J Parallel Distrib. Comput.* **70**, 953–961 (2010)
- Li, Y., Guo, L., Prasad, S.K.: An energy-efficient distributed algorithm for minimum-latency aggregation scheduling in wireless sensor networks. In: International Conference on Distributed Computing Systems, Genova, Italy, pp. 827–836 (2010)
- Xiang, L., Luo, J., Vasilakos, A.: Compressed data aggregation for energy efficient wireless sensor networks. In: 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, San Jose, USA, pp. 46–54 (2011)
- Li, H., Lin, K., Li, K.: Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. *Comput. Commun.* **34**, 591–597 (2011)
- Li, H., Wu, C., Hua, Q.-S., Lau, F.C.M.: Latency-minimizing data aggregation in wireless sensor networks under physical interference model. *Ad Hoc Netw.* **12**, 2014 (2011)
- Kalpakis, K., Dasgupta, K., Namjoshi, P.: Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. *Comput. Networks* **42**, 697–716 (2003)
- Tang, X., Xu, J.: Extending network lifetime for precision-constrained data aggregation in wireless sensor networks. In: IEEE INFOCOM, Spain, pp. 1–12 (2006)
- Chen, I.R., Speer, A.P., Eltoweissy, M.: Adaptive fault-tolerant QoS control algorithms for maximizing system lifetime of query-based wireless sensor networks. *IEEE Trans. Dependable Secur. Comput.* **8**, 161–176 (2011)
- Misra, S., Dias Thomasinou, P.: A simple, least-time, and energy-efficient routing protocol with one-level data aggregation for wireless sensor networks. *J. Syst. Softw.* **83**, 852–860 (2010)
- Dulman, S., Nieberg, T., Wu, J., Havinga, P.: Trade-Off Between Traffic Overhead and Reliability in Multipath Routing for Wireless Sensor Networks. Bernoulli
- Hall, M., Frank, E., Holmes, G., et al.: The WEKA data mining software. *ACM SIGKDD Explor. Newsl.* **11**, 10 (2009)
- Lalka N (2015) Fuzzy Based Expert System for Diabetes Diagnosis and Insulin Dosage Control
- Grosan, C., Abraham, A.: Rule-based expert systems. *Intell. Syst.* 149–185 (2011)
- Orchard R (2002) Fuzzy Reasoning in JESS: The FuzzyJ Toolkit and Fuzzyjess

An Approach Toward More Accurate Forecasts of Air Pollution Levels Through Fog Computing and IoT



Bhavya Deep, Iti Mathur and Nisheeth Joshi

Abstract Rapid urbanization combined with an almost insatiable need for energy has spawned various forms of pollution. Researchers have found air pollution to be at the top of list of factors that cause the most fatalities among urban dwellers today. Scoping urban areas that harbor the most air pollutants and contaminants can help an urban dweller identify comparatively less polluted routes. However, processing of information related to air pollutants is time intensive. As such, temporal forecasting takes preeminence in designing a system that can provide information well in advance of concentration levels of air pollutants at any given time in day or night. In this paper, the authors approach problems related timely forecasts for predicting and tracing air pollution levels across major thoroughfares in urban environments, using fog computing and Internet of Things (IoT). The objective of the research and proposed method is to offer a time-sensitive forecasting to enable citizens to adopt a more agile route-planning approach at any given point of time. In the wake of rising deaths owing to air-borne pollutants and chemicals, results of the research conducted indicate an object-oriented approach toward building a smarter city.

Keywords Fog computing · Internet of things (IoT) · Smart city · Urban pollution · Air quality index · Exponential smoothing

B. Deep (✉)

Bhaskaracharya College of Applied Sciences, University of Delhi, New Delhi, Delhi, India
e-mail: bhavya.deep@bcas.du.ac.in

I. Mathur · N. Joshi

Banasthali Vidyapith, Tonk, Rajasthan, India
e-mail: mathur_itি@rediffmail.com

N. Joshi

e-mail: nisheeth.joshi@rediffmail.com

1 Introduction

Air pollution is arguably the most potent form of killing that has resulted as a consequence of human progress and technological advancements. Pollutants in the air have been identified as one of the leading causes of deaths among humans in urban areas. It has, therefore, grown to be a subject of vital interest to scientists, doctors, researchers, and environmentalists alike to seek ways and means to arrest the number of deaths that occur owing to air pollution alone. It is, without doubt, one of the very major problems that vex researchers today on how to meet this challenge by neither slowing down the pace of technological advancements nor hobbling progressive improvement in quality of our modern day lives that depend on convenience sustained by machines that in turn consume energy.

While there is enormous interest in reducing consumption of energy and make devices more energy efficient, the key problem lies in the fact that a vast bulk of the world's population continues to suffer from the negative effects of energy consumption. Researchers have found that as many as 3.3 million premature deaths are caused by air pollution alone [1]. Even in a country like Germany, air pollution brought about by vehicular traffic accounts for double the mortality rate than that caused by accidents. While there are indeed other serious and equally potent sources of air pollution, which produced from vehicles, is now one of the topmost concerns for urban planners and researchers around the globe.

However, researchers, conducting studies to analyze the impact of finer and more dangerous form of air contaminants with sizes of $2.5 \mu\text{m}$ or less, have concluded that death toll in Asian countries such as China and India could be much higher than in developed nations. This, the researchers have determined, is caused as a result of incomplete combustion of fossil fuels and burning of wood and coal for domestic purposes.

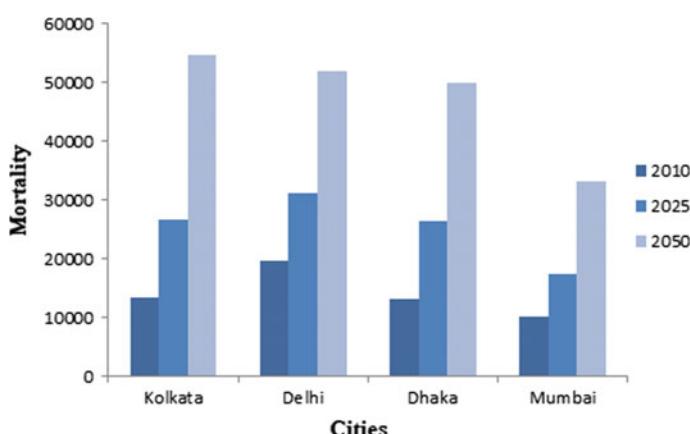


Fig. 1 Mortality linked to air pollution in different cities

It has been estimated that in New Delhi alone, around 30,000 people will have suffered premature deaths as a direct consequence of rising air pollution in 2025. Mumbai, often referred to as the economic capital of India, will also have a high number of premature deaths by year 2025. Kolkata may have a staggering rate of premature deaths that may surpass that of either New Delhi or Mumbai by 2050 [2]. In 2010, New Delhi recorded almost 20,000 premature deaths caused by air pollution. In 2025, the number is expected to be around double of that. In contrast, the cities of London, New York, and Paris may have a combined death toll caused by air pollution of not more than one-fifth of that of New Delhi in 2050. The rates of death tolls in four major South Asian cities are presented in Fig. 1 to reflect the kind of severity that air pollution imposes on the urban populations.

The actual cause leading to premature deaths in urban populations, researchers have observed, is the presence of fine particles carried in polluted air. These particles that can measure no more than $2.5 \mu\text{m}$ (PM2.5) induce wide-ranging health issues involving the heart and lungs of the human body. Deterioration of the heart and lungs is what causes deaths in humans who are exposed to high concentrated levels of air pollution. Unless concrete steps are taken to address the problem of air pollution, the number of premature deaths can grow up to 6.6 million by 2050 [3].

As a classic case of air pollution, the city of New Delhi has no immediate equal. In 2010, the number of premature deaths in New Delhi was recorded at 19700; and the figures are expected to hover above the 30- and 50-thousand mark in 2025 and 2050, respectively [2].

Reports on air pollution in New Delhi [4, 5] discuss and examine two types or sizes of particulate matter that are key to understand this devastating form of death-dealing environment. Air-borne particles that are composed of dust in dry seasons and chemical contaminants that are generated through industrial sources are categorized as PM10. Those that are produced through vehicular traffic involving internal combustion engines fueled by petrol and diesel are categorized and identified as PM2.5. It has been observed that vehicular traffic contributes to as much as 20 and 36% in terms of PM2.5 and NO_x precursor gases, respectively. In winter season, vehicular traffic alone is the second highest contributing source of air pollution in terms of PM2.5 and PM10 [5]. The All India Institute of Medical Science, New Delhi, has reported witnessing a jump of almost 300% in terms of patient admissions related directly to respiratory problems as a consequence of air pollution [6]. What has even been more alarming is the suggestion that concentration levels of PM2.5 particles could be 13 times the number considered safe by the World Health Organization. Alarmed by the present and foreseeable trends in mortality rates, the government and policy makers in New Delhi have decided to roll out the “Smart Cities Mission.” The project is aimed at making at least 100 Indian cities more citizen-friendly and environmentally sustainable in the long run. The planners of this project made it abundantly clear that air pollution figures high in the list of features and priorities of the “Smart Cities Mission” program. Among the eight features [7] that form the core of the mission, reduction in air pollution and diminishing traffic congestion have been included to create environmentally friendly localities for citizens.

Policy makers and urban planners have since taken on board the idea that information and communication technologies (ICT) will need to be the key to support a viable city modernization program. Consequent upon this realization, the “Smart Cities Mission” program espouses the role of mobile-based tools. Like similar other technologies that are reliant on the Internet, mobile-based tools would need to establish a form of communication and effective means of real-time data capture. It is based on this premise that researchers have conducted studies on Internet of Things (IoT) and Cloud of Things (CoT) in the context of urban city planning [8, 9].

Developing on IoT and CoT platforms, we have designed an architecture that would be able to help select the least possible polluted route, or “greenest” possible route, for urban dwellers within a highly congested and high traffic environment. By examining the data available on air pollutants during peak hour periods at different locations in New Delhi, we are able to forecast the levels of air pollutants using exponential smoothing method. By engaging fog computing with IoT, we have designed the system to be able to push air pollution indicator values to end users.

We have organized our manuscript in several sections. We discuss the state of the art in research in our area in Sect. 2. We describe our problem formulation approach and also demonstrate our proposed fog computing-based IoT system infrastructures in Sect. 3. We present our mathematical formulation and, subsequent, result analysis in Sect. 4. Finally, in Sect. 5, we draw our conclusions and discuss future direction that can be made through our findings and proposed system architecture.

2 State of the Art

The number of deaths caused as a consequence of air pollution in our cities and congested urban environments rank just below those caused by heart diseases and smoking [10]. The city of Chicago has devised a system that is referred to as “Array of Things.” This project involves mounting sensors and nodes on lampposts to detect chemical pollutants and particles in the air. So serious is the issue of air pollution that major cities in developed nations consider this to be a serious threat to the health and well-being of its citizens. This “Array of Things” is a significant evolution in terms of IoT and cloud computing paradigms. Desai et al. in their approach to measuring levels of carbon monoxide and carbon monoxide in urban environments have relied on a combination of Global Positioning System (GPS) and Microsoft’s Azure Cloud and Machine Learning services [11]. This is decidedly a stellar example of tagging geographical coordinates with streaming data on air pollution in real time. Indeed, the survey report on air pollution monitoring systems by Khot et al. [12] points to the need to collect live and accurate information on air pollution levels to enable authorities respond as quickly as possible.

However, the need to disseminate accurate and timely reports on present and existing levels of air pollutants in urban regions was acutely felt by Nagarathna et al. [13]. Although the “Smart Cities Mission” program brings to the fore the significance of mobile devices to enable users assess for themselves current levels

of air pollution, the need for automation and increasing the reach among the masses are of vital interest to researchers. In trying to design a forecasting model to support qualitative analysis on PM2.5 concentrations, the authors Liu et al. [14] proposed a comprehensive forecasting model. This model was built using a mix of forecasting ideas involving autoregressive integrated moving average model (ARIMA), artificial neural networks (ANNs) model, and exponential smoothing method (ESM) to predict time series data on PM2.5 concentration levels.

The need to present an accurate forecasting model is the key to prolonging the life and enhancing transmission quality of wireless sensor nodes. Although the authors [15] draw reference to modern wireless sensor hardware that is able to harvest energy and thereby operate autonomously for longer periods, they also point the challenge posed by channel resource limitations in wireless sensors. Against this backdrop, Dias et al. [15] propose a forecasting model that can produce highly accurate forecasting information and, thus, reduce the amount of transmissions by at least 30%. By testing different sensors under varying test conditions, the authors were able to simulate results that demonstrate the efficacy of exponential smoothing method (ESM).

Combining Internet and Web applications with wireless access points (WAP), Roy et al. [16] were able to demonstrate feasibility of delivering real-time data on air pollution levels to users. The authors used Zigbee modules and Zigbee-enabled roadside units to fetch live data on air-borne particles and harmful chemicals. The authors pointed out the relevance of monitoring concentration levels of air-borne harmful substances in urban areas and highly congested traffic intersections. In order to produce accurate and timely information and to reach a wider audience, the authors have designed their architecture based on Zigbee-based system that presents data gathered through Google Maps.

Researchers across the globe are cognizant of the fact that there is an urgent need to construct an efficient system of displaying accurate values on air pollution levels to the general masses in urban regions and other areas where there is a negative impact on environment resulting from factories and power plants. While it has been demonstrated that wireless sensors can be effectively deployed to collect data and present that information in real-time [16] issues involving the life span of the sensors and energy requirements are of paramount interest. Therefore, an architecture that is able to support effective forecasting of data can be of significant use when viewed through real-life considerations such as acquisition costs, maintenance under actual conditions, and accurate presentation of information on air pollution.

3 Problem Formulation

In his research report titled “An Application of Exponential Smoothing Methods to Weather Related Data,” the author Double-Hugh Marera presents his findings concerning exponential smoothing methods [17]. His findings show evidence that simplex exponential smoothing method returns consistently higher accuracy projections insofar as predicting daily weather temperature data is concerned. Exponential

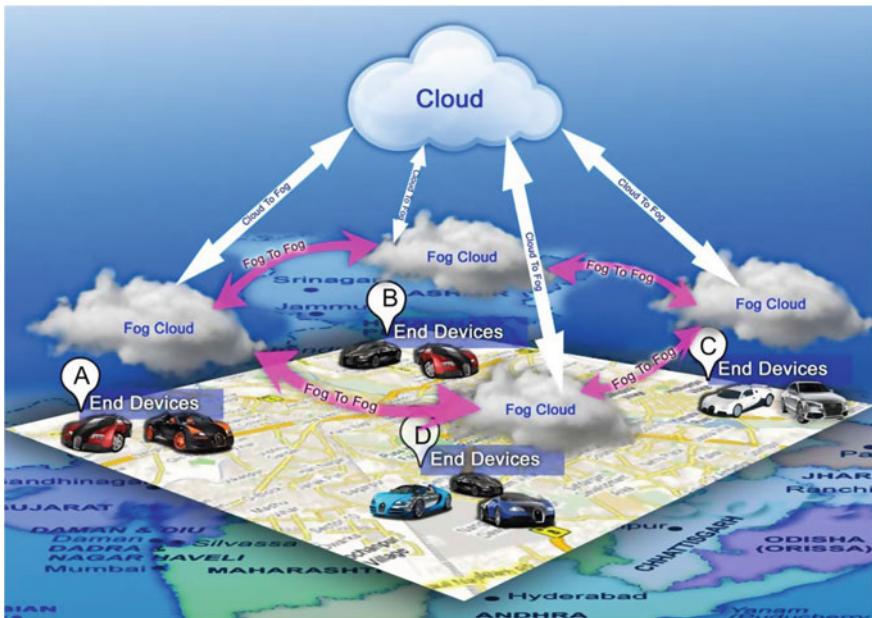


Fig. 2 Fog-based IoT system architecture

smoothing, the author explains, is a technique based on recursive time series. Using this technique, researchers and statisticians are able to update each successive forecast using continuous inputs of real-time data. In simulated test environments conducted, the author observed that simple exponential smoothing method performed the best in cases where forecasts are required just a day in advance. The goal of our research is to strike a balance between computing powers available at wireless sensor level and achieve the objective of providing accurate forecasts to end users in advance. Thus, we aim to accomplish our primary goal of offering forecasts on air pollution levels one day in advance. This would then enable end users to make informed choices on routes with least concentrations of air pollutants.

In this paper, we demonstrate the induction of fog computing to accomplish our goal of working around time intensive and costlier option of uploading data to cloud directly. In doing so, we have shifted processing of data from the core to edge of network(s) [18]. Our proposed design also accommodates the feature of real-time data analysis and uploads data to cloud servers for maintaining historical data storage [19]. An overview of our proposed architecture showing fog-enabled IoT design for generating forecasts based on real-time data is shown in Fig. 2.

4 Mathematical Formula and Result Analysis

Data collected on air pollution levels are considered in time series. Initially, we had collected real-time data on air-borne pollutants in Indira Gandhi International Airport region. The data were manually scraped from the official Web site of Central Pollution Control Board [19]. The data range sets were each of two-hour duration beginning January 15, 2018, to January 18, 2018, and from 8:00 in the morning to 8:00 at night. In order to forecast based on these data sets, we used the single exponential smoothing (SES) formula. The formula [17], for a given time period, is explained as follows:

$$F_t = aC_t + (1 - a)F_{t-1} \quad (1)$$

F_t is called the current smoothed value; C_t is the current observation; and α is called the smoothing constant, where $0 < \alpha < 1$ and $t \geq 3$.

We have expanded on the work involving outdoor air pollutant-monitoring unit (OAPMU). In our proposed fog-enabled IoT system architecture, we envisioned that the Zigbee mounted coordinator module at the traffic junction will perform the data forecasting task and inform to the cloud-based central processing center for different decision-making task. Our forecasts involve five types of data, viz. particulate matter 2.5 ($PM_{2.5}$) and particulate matter 10 (PM_{10}), i.e., 2.5 or 10 μm or less in diameter, ozone (O_3), nitrogen dioxide (NO_2) and carbon monoxide (CO) with $\alpha = 0.1, 0.5$, and 0.9, respectively, which are illustrated in Figs. 3, 4, 5, 6 and 7.

Analyses of the results that are obtained and presented in the graphs that appear in Figs. 3, 4, 5, 6 and 7 reveal that concentration levels of particulate matter and carbon monoxide increase during peak hours. Such concentrations that involve PM_{10} and $PM_{2.5}$ and carbon monoxide are observed to increase in proportion to vehicular movement and density on public thoroughfares and roads at different hours of day and

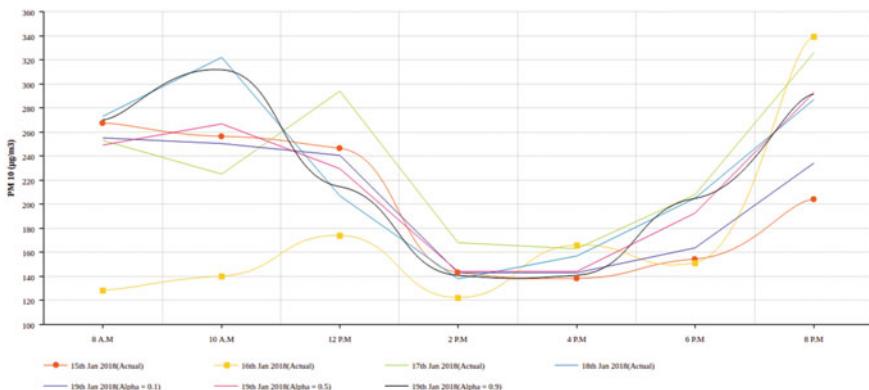


Fig. 3 Forecast with single exponential smoothing on the data of PM_{10}

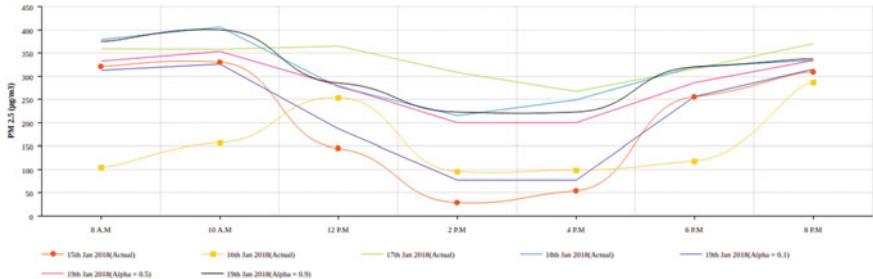


Fig. 4 Forecast with single exponential smoothing on the data of PM2.5

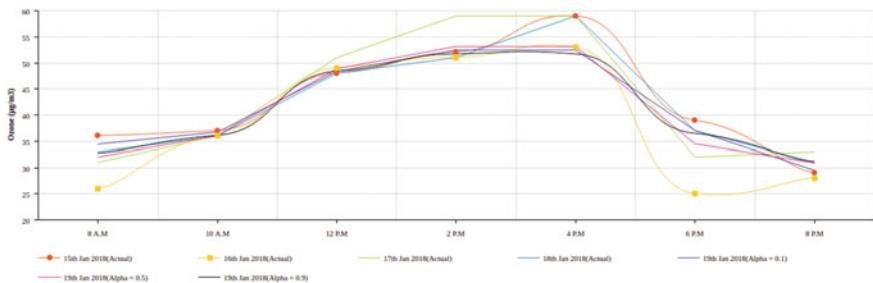


Fig. 5 Forecast with single exponential smoothing on the data of Ozone

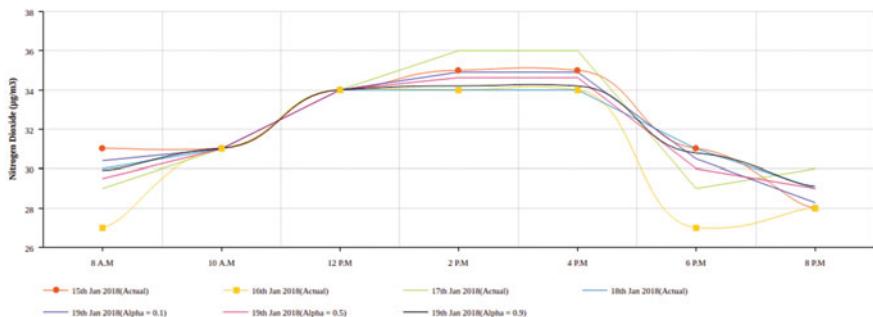


Fig. 6 Forecast with single exponential smoothing on the data of nitrogen dioxide

night. The peak concentration levels are hit within 8:00 and 10:00 in the morning and within 6:00 and 8:00 in the evening hours. During the period 10:00 in the morning till 4:00 in the afternoon, concentration levels of ozone (O_3) and nitrogen dioxide (NO_2) rise. Rise in the levels of ozone and nitrogen dioxide is caused as a result of chemical reactions between oxides of nitrogen (NO_x) with volatile organic compounds (VOC) in the presence of sunlight [16].

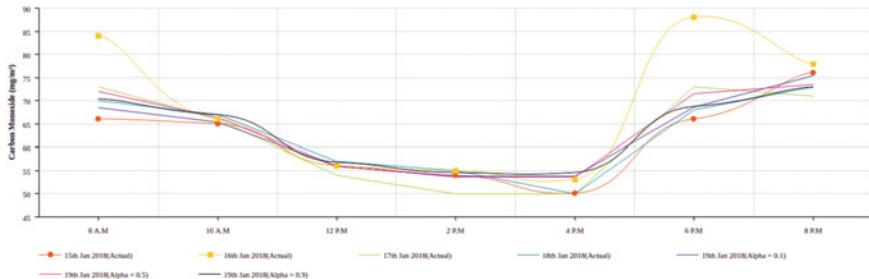


Fig. 7 Forecast with single exponential smoothing on the data of carbon monoxide

5 Conclusion and Future Work

The work that we have presented in our manuscript is significant for two reasons. On one hand, it attempts to provide accurate information related to developing situation concerning air pollutants in an urban environment. On the other hand, it addresses a fundamental problem related to energy conservation and hardware life of the wireless sensors deployed in real-life environments. We propose to mitigate processing loads by introducing fog cloud and IoT to reduce cost and time involved in computing data and uploading it to cloud servers. We have presented an approach whereby urban dwellers can be alerted to air pollution levels at least a day in advance. While this has been possible as a result of forecasting based on real-time data values, forecasting itself extends the hardware life of the sensors that make up the actual IoT framework. Our work can provide a solid base on which further research can be expanded using double and triple exponential smoothing method [17]. Thus far, we have presented a design around exponential smoothing method that is able to produce forecasts for urban dwellers who need to plan travel routes through already congested and highly polluted areas. For urban planners and policy makers, our proposed system is able to offer maintenance benefits in terms of increasing sensor life span. These features can help bolster ideas and boost actual groundwork toward building smart cities of the future. In presenting our proposed design, we have also demonstrated that it is possible to provide a robust framework that is sustainable in the long run and can deliver accurate forecasts within a narrow time frame for the benefit of end users.

References

1. Lelieveld, J., Pozzer, A.: More deaths due to the air pollution—air pollution could claim 6.6 million lives by 2050. Max–Planck–Gesellschaft (2015)
2. Sinhal, K.: Delhi will record world's largest number of premature deaths due to air pollution. TNN (2015)
3. van der Wall, E.E.: Air pollution: 6.6 million premature deaths in 2050! Neth. Heart J. **23**, 557–558 (2015)

4. Delhi Pollution: Dust may be masking bigger killers like vehicular emissions, thermal power plant. First Post (2017)
5. Sharma, M., Dikshit, O.: Comprehensive study on air pollution and green house gases (GHGs) in Delhi. Department of Environment, Government of National Capital Territory of Delhi and Delhi Pollution Control Committee, pp. 1–289 (2016)
6. Karnik, M.: In 2025, Delhi's air will be the world's deadliest – killing over 30,000. Quartz India (2015)
7. Smart Cities Mission: <http://smartcities.gov.in/content/innerpage/what-is-smart-city.php>
8. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: Internet of things for smart cities. IEEE Internet Things J. **1**, 22–32 (2014)
9. Roy, S., Sarddar, D.: The role of cloud of things in smart cities. Int. J. Comp. Sci. Info Secur. **14**, 683–698 (2016)
10. Clark, J.: Big data, pollution, and the IoT. Internet of Things Blog. Environment (2017)
11. Desai, N.S., Alex, J.S.R.: IoT based air pollution monitoring and predictor system on Beagle bone black. In: ICNETS2, Chennai, India (2017)
12. Khot, R., Chitre, V.: Survey on air pollution monitoring systems. In: ICIIECS, Coimbatore, India (2017)
13. Nagarathna, R., Manoranjani, R.: An intelligent step to effective e-governance in india through e-learning via social networks. In: MITE, Madurai, India (2017)
14. Liu, D-J., Li, L.: Application study of comprehensive forecasting model based on entropy weighting method on trend of PM2.5 concentration in Guangzhou, China. Int. J. Environ. Res. Public Health **12**, 7085–7099 (2015)
15. Dias, G.M., Bellalta, B., Oechsner, S.: On the importance and feasibility of forecasting data in sensors. [arXiv:1604.01275v1](https://arxiv.org/abs/1604.01275v1) [cs.NI] 1–30 (2016)
16. Roy, S., Bose, R., Sarddar, D.: Smart and healthy city protecting from carcinogenic pollutants. Int. J. Appl. Environ. Sci. **12**, 1661–1692 (2017)
17. Marera, D-H., Bechelt, F.: An application of exponential smoothing methods to weather related data. Research Report, 1–103 (2016)
18. Roy, S., Bose, R., Sarddar, D.: A fog-based DSS model for driving rule violation monitoring framework on the internet of things. IJAST **82**, 23–32 (2015)
19. Air Quality Data: CPCB official website, <http://cpcb.nic.in/>

Implementation of Dalal and Triggs Algorithm to Detect and Track Human and Non-Human Classifications by Using Histogram-Oriented Gradient Approach



Vijayalakshmi S. Katti, Sushitha S, Shweta Dhareshwar and K. Sowmya

Abstract In this paper, the proposed application is mainly used for security purpose and it deals with searching based on the given input and integrates this approach with the concept of histogram of oriented gradient (HOG) features to achieve person or non-person classification in human tracking system. It captures the features of human automatically based on the gradients of an image in human tracking system. This tracking system process will be done only through the expert team. Before getting into the process, the images or videos which have been captured from the surveillance camera have to be uploaded and checked whether it is a human being or not.

Two leading approaches to the application are:

- Single-detection window approach
- Part-based approach.

This application process is done based on some scenarios like color code, distance between the object and the camera, and the height of the human being.

Keywords K histogram of oriented gradients (HOGs) · Support vector machine (SVM) · Human tracking system (HTS) · Common type system (CTS)

V. S. Katti · S. S · S. Dhareshwar · K. Sowmya (✉)

Department of MCA, Nitte Meenakshi Institute of Technology, Bengaluru, Karnataka, India
e-mail: sowmya.k@nmit.ac.in

V. S. Katti
e-mail: viju.s.katti@gmail.com

S. S
e-mail: sushitha.s@nmit.ac.in

S. Dhareshwar
e-mail: shwetagnmit.2014@gmail.com

1 Introduction

The application used in detecting humans based on the given input and integrates with the concept of Histograms of Oriented Gradients (HOG).

Scenario 1: **Color**—It selects the color based on RGB, gray scale, square root, and log.

Scenario 2: **Distance between the Object and the surveillance camera**—It calculates the distance between the object and the camera through focal length, object height in pixels, and image height in pixel.

Scenario 3: **Based on the height of the human being**—It calculates the height of the human based on sensor value, focal length, and distance between cameras.

2 Statement of the Problem

The aim of the application is to track human motion in image sequences. Pedestrian detection can be achieved using real-time images with low resolution. Detecting people in images is a very difficult problem because of some different poses. It is a main challenge in human tracking system where people dress might be in unique colors or it might be different depending upon the situation, and thus, it is very difficult to define a single-detection approach that describes all the people. People can be walking or running in a real-time scenario, but to get the appropriate information of real-time images, we need to capture side view or slightly rotated view of a person, which changes the outline shape of a body.

3 Related Analysis

The present system detects only the face in real-time images.

Detection is done on the three features: (1) integral image representation, (2) AdaBoost algorithm, and (3) cascade.

Face detection application also plays an important role because the recognition of face is the first step when we want to identify whether the given input is human or not. In some cases, by seeing an image we can easily rectify it is a human, but in most of the time, the challenges faced are articulated poses in variety. In face detection, detecting the color of the skin is the most important technique and powerful which is been done based on the red, green, blue (RGB) format using MATLAB functions. In this application, the main drawback is different poses or positions, occlusions, and the face which is been detected might be object (i.e., dummy).

4 Current Work

The proposed application detects humans in real-time images and in a video which is been captured. The image or video which is been captured will be in low resolution. Most of the time, images which are uploaded will be a static image with a normal background.

The key challenges of the application are:

- Variety of articulated poses.
- Occlusions.
- Background of the image will be complex.
- Illumination will be unconstrained.

This application is mainly used for security purpose. It helps us to track the human based on the input which is given by a particular user or admin. The first and very challenging task in the proposed application is finding whether human or an object from the image or video which is uploaded. Human identification or human recognition is done on the concept of histogram of oriented gradients (HOGs). In HOG image which is been uploaded will convert to gradient image and with the orientation of cells and that cells will be put into blocks over normalization and collect HOG on window. Finally, we can classify it as a human or an object. Object detection concept is been used in video sequence for the subtraction of background regions or color. Detection of human framework is very fast based on the concept like integral HOG and variable size of blocks Training the cascade.

5 System Specification

Minimum Hardware and Software requirement for using the application

Processor—Intel(R) Core i3 processor 3.20 GHz

RAM—512 MB or higher

Hard disk—10 GB or higher

Camera—Webcam 2 MP or higher

Framework—.NET framework 4.5

OS—Windows 7 or higher.

6 System Architecture

Figure 1 represents the System Architecture where the input image which is been uploaded will convert to gradient image and with the orientation of cells and that cells will be put into blocks over normalization and collect HOG on window using linear SVM. Finally, we can classify it as a human or an object.

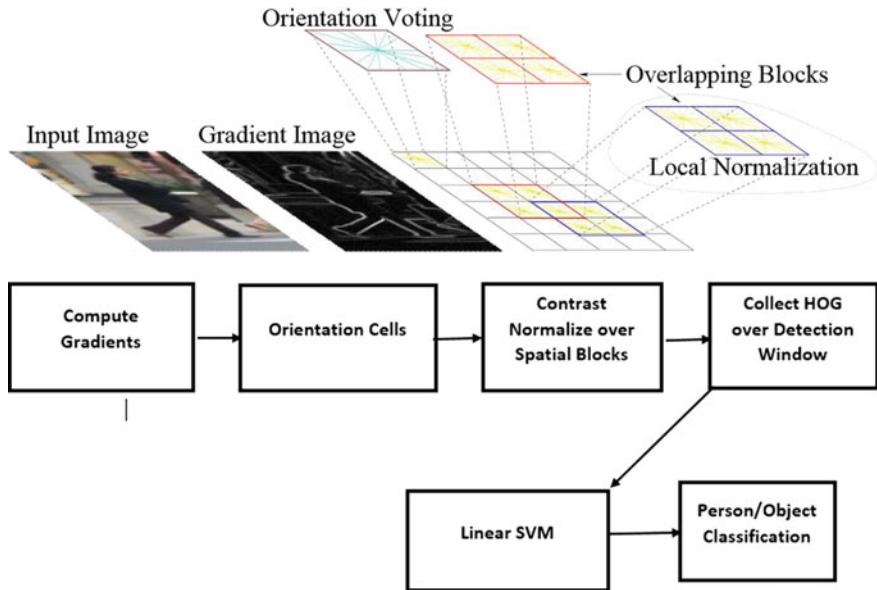


Fig. 1 System architecture

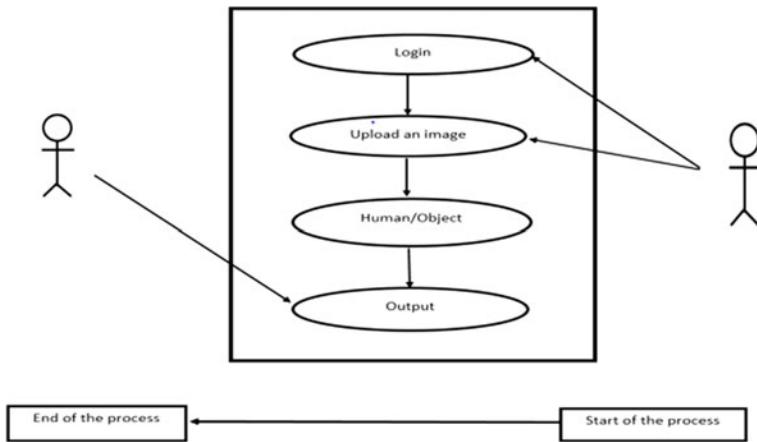


Fig. 2 Usecase diagram

7 Use Case Diagram

See Fig. 2.

8 Flowchart

See Fig. 3.

9 Implementation

Implementation environment

Detection of human tracking is a challenging task because of the huge variety in the appearance of human due to clothing, occlusion, and illumination-constrained conditions that are common in outdoor sequences. The algorithm which is undergone in this process is Dalal–Triggs algorithm. The algorithm is presented with excellent detection results, and the algorithm uses an approach of histogram of oriented gradients (HOGs). It captures the features of human automatically based on the gradients of an image using linear SVM.

Two leading approaches of human tracking system are:

- **Single-detection window approach**

It uses single framework for detecting humans which is been uploaded in the video or image format; i.e., humans will be detected in single-window approach which is shown in the interface design, and usually, the background of the image will be very clear (road, wall, etc.)

- **Part-based approach**

This approach, as the name implies each part of human will be detected separately and if some or all of its parts are presented in proper manner.

The major three components of Dalal–Triggs algorithm:

- HOG is used as a basic building block.
- Dense grid of HOG provides a description of the detection window.
- Normalization is done in every block to emphasize the behavior in connecting to the linear support vector machine with respect to the neighboring cells.

10 Modules Description

A. *Scenario 1: Color*

- Gamma/color normalization

Several input pictorial representations were evaluated using gray scale, RGB, and LAB color spaces optionally with gamma equalization. These normalizations have only a finest effect on performance and thus achieve similar results. We do use the

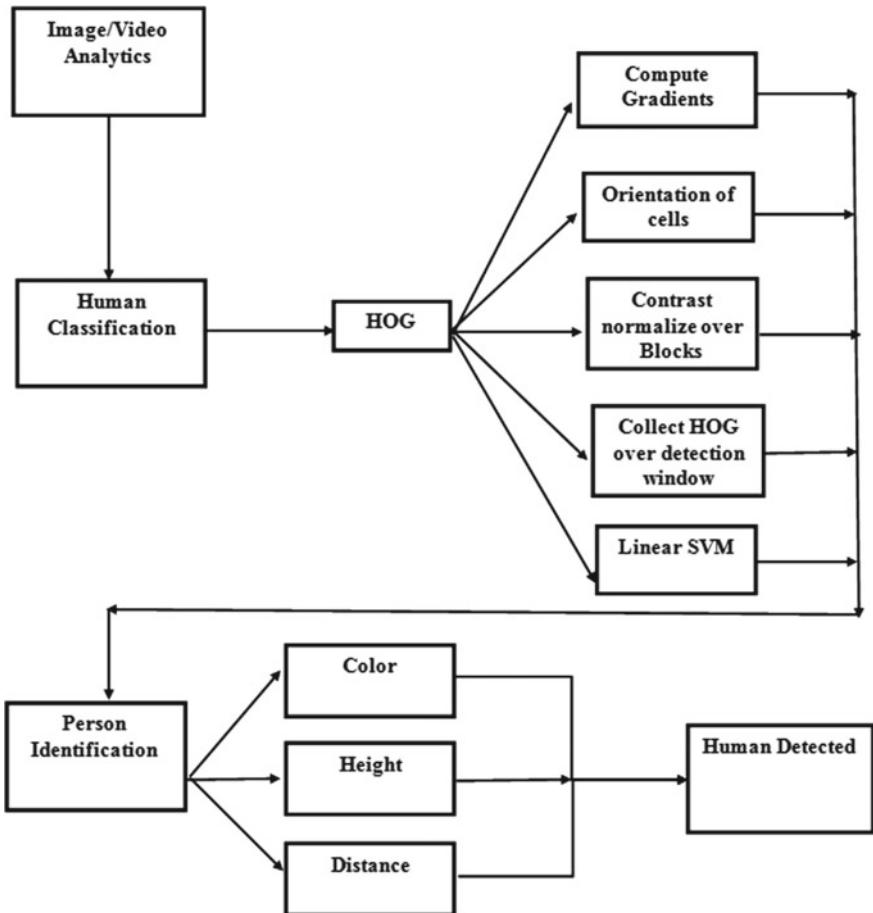


Fig. 3 Flowchart diagram

color information whenever it is available. RGB and the LAB color spaces give comparable results, but restricting to gray scale reduces performance. Square root and log give information about normalization and gamma compression of each color channel to improve performance.

B. Scenario 2: Distance Between the Object and the Camera is Calculated based on

- Focal length (mm)
- Image height (in pixels)
- Sensor value.

The focal length of an optical system is a measure of a system that is how the light will be converged from one point and diverged to another point. To calculate the

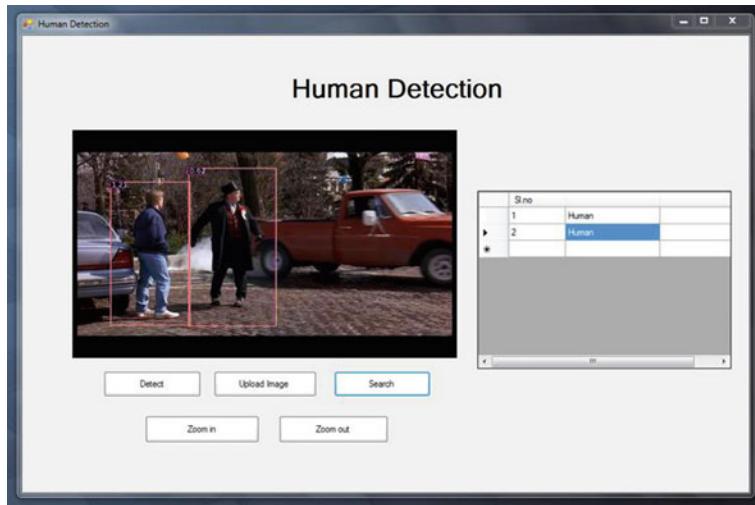


Fig. 4 Human detection screenshot

focal length of an optical system in the air, it calculates the distance in which the rays are brought into a focus. The focal length which has shorter length has the high optical power than the rays bend the light more strongly and brings them to a focus in a shorter distance.

C. Scenario 3: Height of the human being is calculated based on

- Distance between the camera Focal length
- Sensor value.

Every sensor is designed to work over a specified range. The ranges or values are usually assigned, and if we want the absolute output, the sensor value should be more accurate based on the application, and each sensor value varies a lot. It is more efficient to use the elements over only the part of their sensor value and enhanced linear support.

11 Results

A. Human detection screenshot

Figure 4 shows the result of human detection screenshot, where as soon as image is been uploaded when we click on detect and search button, it will be identified as a human/object.

B. Part-based approach screenshot

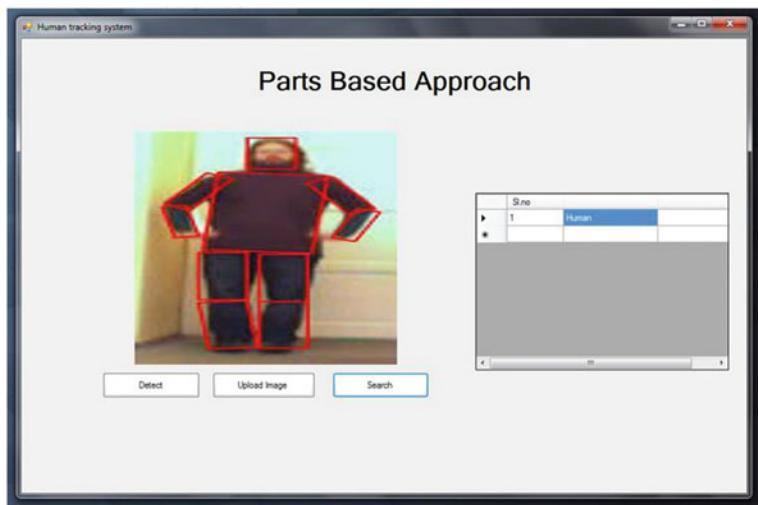


Fig. 5 Part-based approach screenshot

Figure 5 shows the result of part-based approach screenshot where it identifies human/object based on the shape.

C. Color identification screenshots

Figure 6 shows the result of the color identification screenshots that were based on the color it identifies and displays.

D. Distance calculation screenshot

Figure 7 shows the result of distance calculation screenshot where it calculates the distance of human from the camera.

E. Face count screenshot

Figure 8 shows the result of face count screenshot where it calculates how many faces are present in the given input image.

E. Video analysis screenshot

Figure 9 shows the result of video analysis screenshot where it analyzes the captured video.



Fig. 6 Color identification screenshots

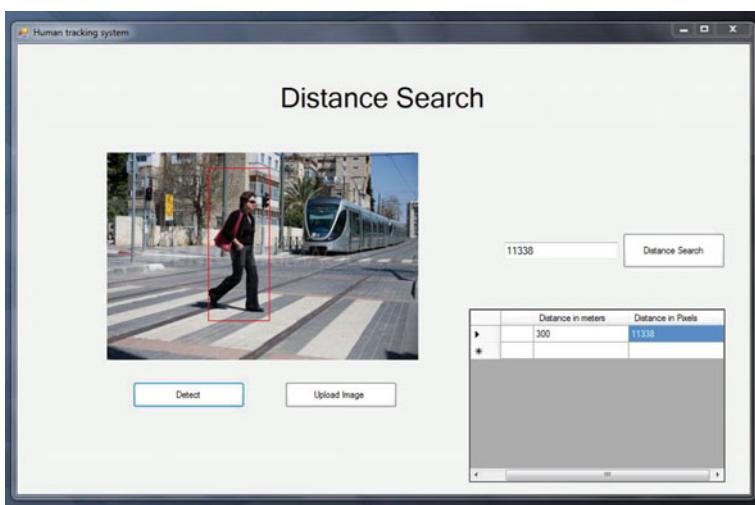


Fig. 7 Distance calculation screenshot

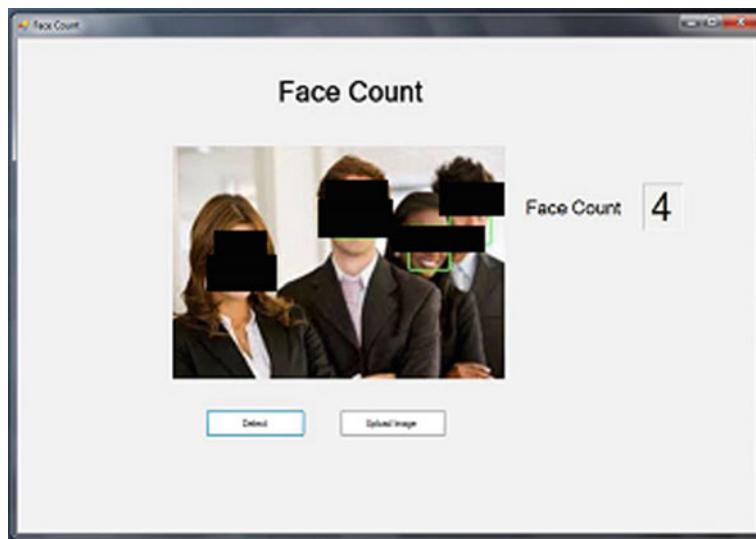


Fig. 8 Face count screenshot

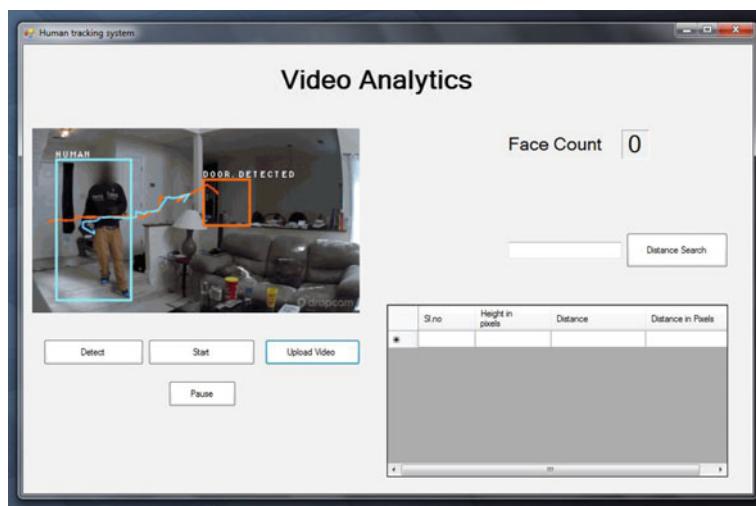


Fig. 9 Video analysis screenshot

12 Conclusion

We demonstrate that human tracking system tracks the human and helps us to find out whether human or non-human/object in real-time images or videos based on the concept of histogram of oriented gradients (HOGs). This HTS will help to find the color of the human dress and the height of the human being and also calculates the distance from the object which is uploaded from surveillance camera based on the input. It also counts the faces in the real-time images and also converts the video format into images which can be helpful for detecting the human/object.

13 Future Enhancements

User will keep on varying the requirements, so it is not possible to fulfill all the requirements of the user.

Some of the enhancement can be done:

- In human tracking system, we cannot find the gender of humans, so based on that, experiments can be done and efficient algorithms can be used to find the solution.
- In this system, we do not use any kinds of the database to store or retrieve, so keeping that in mind, new database concepts can be adapted to this application.
- Security-wise can be improved by giving alert signals.

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of IEEE International Conference on Computer Vision Pattern Recognition, pp. 886–893 (2005)
2. Ye, O., Han, Z., Jiao, J., Liu, J.: Human detection in images via piecewise linear support vector. *IEEE Trans. Image Process.* **22**(2) (2013)
3. Pang, Y., Yan, H., Wang, K.: Robust CoHOG feature extraction in human centered image/video management system. *IEEE Trans. Syst. Manage. Cybern.* **42**(2), 458–468 (2012)
4. Jun, B., Choi, I., Kim, D.: Local transform Features and hybridization for accurate face and human detection. *IEEE Trans. Patt. Anal. Mach. Intell.* **35**(6), 1423–1436
5. Zhu, Q., Avidan, S., Yeh, M., Cheng, K.: Fast human detection using a cascade of histograms of oriented gradients. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2006). Xu, F., Gao, M.: Human detection and tracking based on HOG and particle filter. *IEEE International Congress on Image and Signal Processing (CISP-2010)*, pp. 1503–1507 (2010)
6. Lee, K., Choo, C.Y., See, H.Q., Lee, Y.: Human detection using histogram of oriented gradients and human body ratio estimation. *IEEE Conference on Computer Vision*, pp. 18–22 (2010)
7. Xu, R., Zhang, B., Ye, Q., Jiao, J.: Cascaded L1-norm minimization learning (CLML) classifier for human detection. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 89–96 (2010)
8. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. *Int. J. Comput. Vis.* **63**(2), 153–161 (2005)
9. Oladunni, O., Singhal, G.: Piecewise multi-classification support vector machines. In: Proceedings of International Joint Conference on Neural Network, pp. 2323–2330 (2009)
10. Ren, S.Q., Yang, D., Li, X., Zhuang, Z.W.: Piecewise support vector machines. *Chin. J. Comput.* **32**(1), 77–85 (2009)
11. Cheng, H.B., Tan, P.-N., Jin, R.: Efficient algorithm for localized support vector machine. *IEEE Trans. Knowl. Data Eng.* **22**(4), 537–549 (2010)
12. Ge, J., Luo, Y., Tei, G.: Real time pedestrian detection and tracking at nighttime for driving assistance system. *IEEE Trans. Intell. Transport. Syst.* **2**, 283–298 (2009)
13. El-Naqa, E., Yang, Y., Werick, M.N., Galatsanos, N.P., Nishikawa, R.M.: A support vector machine approach for detection of micro-calcification. *IEEE Trans. Med. Imag.* **21**(12), 1552–1563
14. Ammar, B., Wali, A., Alimi, A.M.: Incremental learning approach for human detection and tracking. *IEEE International Conference on Innovations in Information Technology*, pp. 128–133 (2008)

Safe Navigation for Elderly and Visually Impaired People Using Adhesive Tactile Walking Surface Indicators in Home Environment



Rajanala Vijaya Prakash and Srinath Taduri

Abstract Elderly and visually impaired people will have low vision due to aging problem, disease, or accident. These people encounter difficulties in conducting an independent life, which is intrinsic to the nature of their impairment. Navigation in the known/unknown environment seems difficult without external help. There is a need of self-explanatory tactile indicators sensing through their foot and navigate in a home environment. The home environment layouts are dynamic in nature as furniture and other items' position will be changed frequently. In this paper, we designed and developed an adhesive tactile walking surface indicators (ATWSI) and tactiles to navigate in the home environment with the help of a 3D printer technology using acrylonitrile butadiene styrene (ABS) and thermoplastic polyurethane (TPU) material.

Keywords Tactile · Navigation · Adhesive · Impaired · 3D printer · ABS · TPU

1 Introduction

Aging will lead to many health problems; the visual disorder is one of the common and serious problems. These may be cataracts, glaucoma, macular degeneration, and retinal complications from diabetes. These diseases change the visual effects of a person like blurring, partial loss of visual field, through to genuine visual hallucinations, and complete blindness. Aging people will regularly take some medications; the medications can also lead to visual side effects [1, 2].

Elderly with low vision and visually impaired people face discreet levels difficulty in day-to-day life to accomplish their tasks like eating or moving from place to place in their environment or habitat. These problems get differentiated from person to person based on their target of work. In this progressive world, each and every problem can be solved with either conventional or unconventional thinking and the

R. Vijaya Prakash (✉) · S. Taduri
S R Engineering College, Warangal, India
e-mail: vijprak.r@gmail.com

quality of life depends on how the proposed solution is acceptable to most of the community. For this problem, there is a need for a solution where visually impaired people will be capable of taking decisions and moving around on their own [3].

Visually impaired people will either move around independently or with the aid of a sighted person who will act as a guide. Those who move around independently will do so either solely by using their residual sight or by using mobility aid [2, 4, 5]. The most common mobility aid used by pedestrians with poor sight to facilitate their independent mobility is a long white cane [2, 6]. This is used to scan the ground in front of the person. The scanning takes the form of sweeping the cane in an arc from one side to the other to just beyond the width of the body [2]. This technique will locate potential obstructions such as street furniture, provided that there is some element at ground level, and distinct changes in level such as a kerb upstand or a step. In their own home environment, elderly and visually impaired people must exhibit efforts and cognitive burden to navigate from one place to another avoiding obstacle.

Let us suppose a blind person is walking into a room and he needs a set of task to be accomplished like picking up an object which is lying on a table or like moving into a lobby space to sit and relax for a while. First and the foremost thing that gets into his mind are where to start and how to start? And let us hope somehow, he started and the next question is how long the object he needs? What are the obstacles in between to cross the distance without getting hurt or by avoiding obstacles? In the room, there are many obstacles like chairs or furniture will be placed in random position and their position is not predefined and it would be hard for a visually impaired person to identify whether to take right or left or move forward or go back. To overcome this kind of situation, there is a need for an assistance system so that it would be easy for people and help them reach their destination.

Tactile paving surfaces can be used to convey important information to visually impaired pedestrians about their environment, for example, hazard warning, directional guidance, or the presence of an amenity [7–9]. Each type of tactile paving surface should be exclusively reserved for its intended use and consistently installed in accordance with these guidelines. Visually impaired people are becoming increasingly mobile, both within their local area and more widely, and it is, therefore, very important that conflicting and confusing information is not conveyed. These tatics are made of ceramic-based and once laid it on the ground or area it cannot be reused. Another aspect of these tactiles is it will be used for the outdoor environment. When the navigation environment obstacles are changing dynamically like changing the furniture layouts in the home environment then there is a requirement of adhesive tactile, which can be easily rearranged.

For these reasons, we are proposing an adhesive tactile walking surface indicator tactile for elderly and visually impaired people for easy navigation in a home environment. These adhesive tactile tiles are flexible and made with a reusable plastic material with the help of 3D printer technology using ABS and TPU materials.

2 Related Work

Tactile walking surface indicators are widely used for navigation in an outdoor environment. The design is based on Canadian (CSA) and International (ISO) standards [10, 11]. These tatics are first used by Japan since 1967. The ISO standard tactile is widely used for blind people for safe navigation in outdoor environments especially in the pathways [1, 2, 7, 12]. The Canadian National Institute for the Blind has also published a guide to design the tactile for blind in 2009 [5, 12, 13]. The Government of Ontario also had a significant consultation process involving advisory groups and representatives of various disabilities in the development of its accessibility standards [12]. In addition, a pilot project was conducted by the city at an intersection by St. Michael's Hospital from November 2012 to July 2013 [10]. A Disability Issues Committee is constructed by St. Michel's Hospital, and this committee tested the various issues related to the navigation like public consultation, media interest, seniors, persons with mobility issues and mobility instructors for the visually impaired [10, 13].

Based on these standards and inputs, researchers are designing and developing assistive technology for the elderly and visually impaired people, but still, it is a challenging and ambitious research area [5]. Most of these assistive technologies are used in outdoor navigation using ceramic tile for pathways, mobile, sensor-based white cane or mobile app-based navigation [10, 14–16]. The white cane can be used for the outdoor environment; it cannot be used extensively in the indoor or home environment. Elderly one cannot use the mobile and sensor-based assistive technologies due to unaware of these technologies.

The simple and easy tool for safe navigation for visually impaired people is using a tactile. These tactile tiles are ceramic-based and especially used for the outside environment in the pathways [17]. The ceramic-based tactile are static, once it is laid it cannot be changed. For these reasons, there is a need for the adhesive tactile walking surface indicator. These tactile can be easily rearranged as per the demand and requirement in the home environment for safe navigation by elderly and visually impaired people.

3 Designing the System

The system has three major phases: understand the design parameters; designing the tactile; and test the tactile on the target group.

The design parameter of the visual indicator layer involves understanding the tactile and visual capabilities of the target customer group. Limited research is available on tactile capabilities of the feet of elderly people. This area needs additional research as the design of tactile requires the elderly to feel the product through their feet.

The experiments will be carried out in a darkroom mounted with projectors. The initial experiments use projectors to project to image in front of the participant on

the wall different colors, color contrasts, and patterns. The images could be words or objects, which incorporate colors/color contrasts/patterns. The participants are asked to

1. Evaluate whether they could see the design
2. Trigger a switch to determine the response time to see the image.
3. Describe their feeling including whether the image was too bright or too dim, the contrast
4. Level and ease of seeing the pattern.

To determine the tactile abilities, the participants are asked to walk slowly moving the feet while in gentle contact with the floor. Different adhesive strips with various thicknesses and texture are taped on the floor. The participants are asked to

1. Evaluate whether they felt the tape
2. Describe their feeling including whether the tape was too thick or too thin possibility of tipping, comfort level, and thermal profile.

The texturing layer is a crucial problem in the design of this experiment. Texturing will be done using thin 3D printed strips which are bonded to the duct tape. Thus, these experiments provide the quantitative feedback of effective tactile indicators for easy navigation in a home environment for elderly and visually impaired people.

Based on the feedback and inputs from the target group, the following tactile walking surface indicators are identified to design the tactile (Fig. 1).

Using these tactile indicators, there are three tiles are designed as specified in Fig. 2 namely warning tile, directional tile, and turn tile. The warning tile is used when any hazards like wall or furniture occur in the navigation. This indicates that the user must stop his/her navigation. When the user senses the texture available in the directional tile, he/she can go forward or navigate in the home. The turn tile is placed when the path in the home environment requires either the left side or right side. The targeted user will sense the tile with his/her foot receive the feedback, based on the feedback received from the tile, the user will decide the turn and move accordingly.

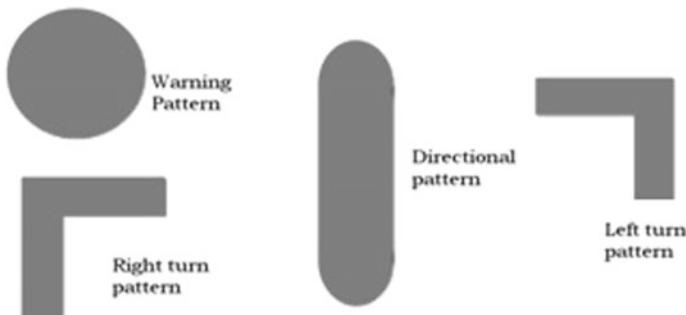


Fig. 1 Tactile walking surface indicator

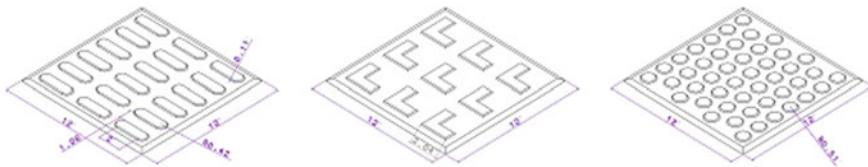


Fig. 2 Directional, turn, and warning tiles

4 Experiments

The experiment is conducted on elderly and visually impaired people. Elderly people are classified into three categories: moderate users, severe users, and profound users based on their vision. Each user is asked to identify the indicator through their foot. Then they are asked to walk on these tiles to make safe navigation in the indoor environment. The flow of this procedure is explained with a flow chart specified in Fig. 3.

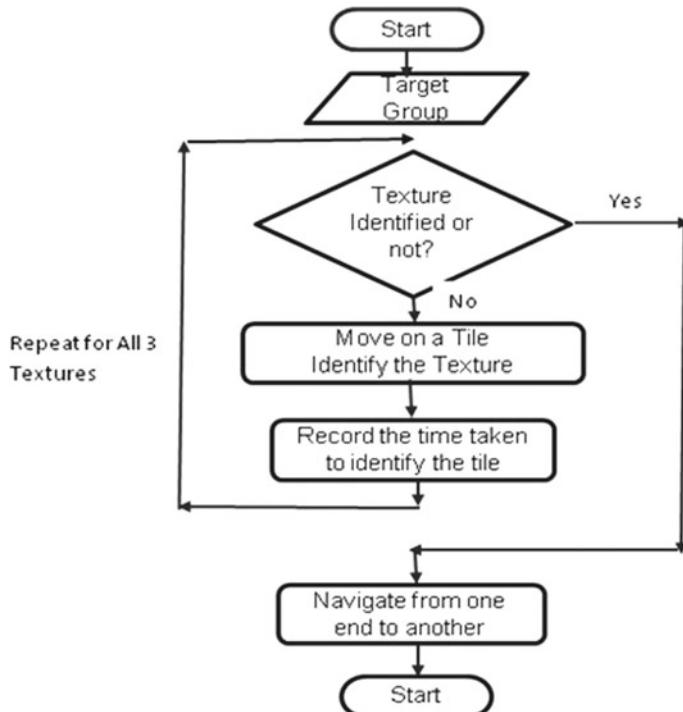


Fig. 3 Flow chart for navigation

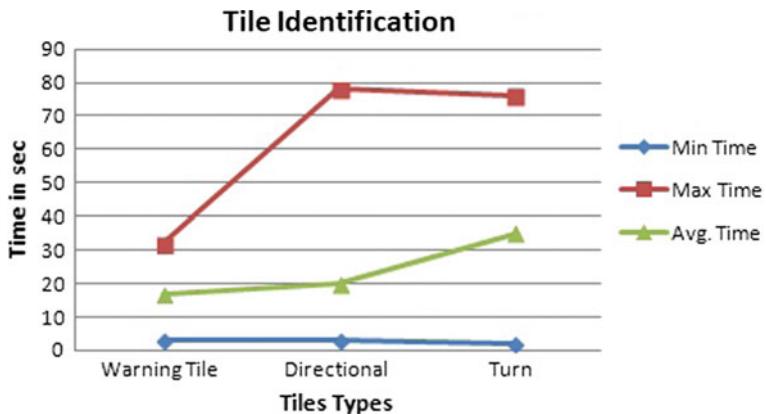


Fig. 4 ATWSI tile identification

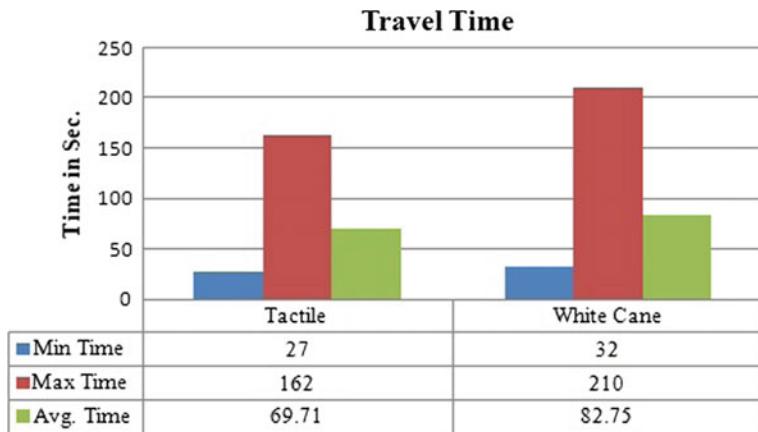


Fig. 5 Navigation

The tactile indicator identification time in seconds is recorded. It is observed that the warning indicators are quickly identified by the users and other indicators are taken some more time compared to warning tactile indicator. The identification times of tactiles are in the order of warning indicator, directional, and turn tiles; these are specified in Fig. 4.

After the identification of tactile indicators, we also recorded the navigation time taken by the different users using the tactile and white cane. The average time taken to navigate safely in the home environment on different paths is specified in Fig. 5.

5 Conclusion

This paper presents a safe navigation system for elderly and visually impaired people in the home environment. The system provides self-explanatory tactile indicators, which are sensed by their foot. Based on this sensing knowledge, the user can easily navigate in an indoor environment. The home environment layouts are frequently modified, thus it is essential to have an adhesive-based tactile. People those who are living in the rented houses, they frequently/occasionally shift from one place to another, then their home environment will be changed accordingly. For these scenarios, the adhesive tactile can be easily rearranged for safe navigation by elderly and visually impaired people. The proposed tactile are designed and developed with the help of 3D printing technology using ABS and TPU materials.

Acknowledgements The authors gratefully acknowledge the helpful comments, suggestions of the reviewers, and continuous support from Management, Principal of S R Engineering College, Warangal Urban, which have improved the quality of this paper. This research is supported by the Department of Science and Technology, India (SEED/TIDE/035/2015).

Ethical Statement

Written informed consent was obtained from all the participants. The protocol for this study was approved by the Institutional Review Committee at SR Engineering College, Warangal.

References

1. Brouwer, D.M., Sadlo, G., Winding, K., Hanneman, M.I.G.: Limitation in mobility: experiences of visually impaired older people. *Br. J. Occup. Therapy* **71**(10), 414–421 (2008)
2. Desrosiers, J., Wanet-Defalque, M.-C., T'émisjian, K., Gresset, J., Dubois, M.-F., Renaud, J., Vincent, C., Rousseau, J., Carignan, M., Overbury, O.: Participation in daily activities and social roles of older adults with visual impairment. *Disabil. Rehabil.* **31**(15), 1227–1234 (2009)
3. Eizmendi, G., Azkotia, J.M., Craddock, G.M.: Challenges for assistive technology. In: AAATE 07, vol. 20. Ios Press (2007)
4. Nyman, S.R., Dibb, B., Victor, C.R., Gosney, M.A.: Emotional well-being and adjustment to vision loss in later life: a meta-synthesis of qualitative studies. *Disabil. Rehabil.* **34**(12), 971–981 (2012)
5. Steel, E.J., de Witte, L.P.: Advances in European assistive technology service delivery and recommendations for further improvement. *Technol. Disabil.* **23**(3), 131–138 (2011)
6. HumanWare, “Trekker Breeze”: http://www.humanware.com/en-usa/products/blindness/talking_gps/trekker_breeze/details/id_101/trekker_breeze_handheld_talking_gps.html. Accessed 21 Mar 2016
7. Demirkan, H.: Effectiveness of tactile surface indicators in design for all context. *Open House Int.* **38**(1) (2013)
8. Emma, J., Agneta, S., Susanne, I., Synneve, D.I., Mai, A.: Blind people testing the tactility of different surfaces in experimental and real environments. In: 12th International Conference on Mobility and Transport for Elderly and Disabled Persons (TRANSED) (2010)
9. Qvistdal, L.R., Lid, I.M., Lindland, T.: How to evaluate the effectiveness of a tactile surface indicator system. In: International Congress Series, vol. 1282, pp. 1051–1055. Elsevier, Amsterdam (2005)

10. Pitt, I., Edwards, A.: Improving the usability of speech-based interfaces for blind users. In: International ACM Conference on Assistive Technologies, New York, NY, USA, pp. 124–130. ACM (1996)
11. Espinosa, M.A., Ochaita, E.: Using tactile maps to improve the practical spatial knowledge of adults who are blind. *J. Vis. Impair. Blind.* **92**(5), 338–345 (1998)
12. Bentzen, B.L., Barlow, J.M., Tabor, L.S.: Detectable warnings: synthesis of US and international practice, pp. 17–21. US Access Board (2000)
13. Innala Ahlmark, D., Fredriksson, H., Hyypä, K.: Obstacle avoidance using haptics and a laser range finder. In: 2013 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 76–81 (2013)
14. Hakobyan, L., Lumsden, J., O'Sullivan, D., Bartlett, H.: Mobile assistive technologies for the visually impaired. *Surv. Ophthalmol.* **58**(6), 513–528 (2013)
15. Guerrero, L.A., Vasquez, F., Ochoa, S.F.: An indoor navigation system for the visually impaired. *Sensors* **12**(6), 8236–8258 (2012)
16. Sound Foresight Technology: Ultracane—putting the world at your fingertips. <http://www.ultracane.com/>. Accessed 21 Mar 2016
17. Kobayashi, Y., Osaka, R., Hara, T., Fujimoto, H.: How accurately people can discriminate the differences of floor materials with various elasticities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **16**(1), 99–105 (2008)

Extensible Attribute Similarity Data Mining for Categorical Data Streams in Web Usage Framework



N. Pushpalatha, S. Sai Satyanarayana Reddy and N. Subhash Chandra

Abstract Information mining is a forcefully idea in data recovery in view of various characteristics from various information sources. Emergence increase of web 2.0 oriented applications for best user encounters and availability without time and geographical limitations. For the researchers, Web usage logs are becoming a major role across the world, and user's behavior and analysis of data are one of the different concepts for decision making in business intelligence. To employ customer-centric Web-oriented applications, we implement extensible Web usage mining framework (XWUMF), i.e., hybrid framework to handle data extraction based on user behavior with different attributes in data sets. Proposed hybrid approach is the combination of extensible and classification by pattern-based hierarchical clustering (ECPBHC). Extensible and classification by pattern-based hierarchical clustering (ECPBHC) is used to extract relational data from different Web-oriented categorical data sets based on user's behavior and analysis with different attributes' relations.

Keywords Data mining · Hierarchical clustering · Fuzzy clustering · Similarity index and support vector machine

1 Introduction

Information mining is a forceful idea in data recovery in view of various traits from various information sources. For powerful information accumulation from infor-

N. Pushpalatha

Marri Lakshman Reddy Institute of Technology and Management, Hyderabad, India
e-mail: pushpalatha523@gmail.com

S. Sai Satyanarayana Reddy (✉)

Vardhaman College of Engineering, Hyderabad, India
e-mail: saisn90@gmail.com

N. Subhash Chandra

Holy Mary Institute of Technology, Hyderabad, India
e-mail: subhashchandran@yahoo.co.in

mation sources regarding important information, one-class learning is required to perform mark-based order with singular preparing successions on qualities. For some true information outsourcing constant informational collection assigning with unusual social class mark occasions with costly unthinkable information introduction. To take in these kinds of aggregate successions continuously informational index procedures to characterize target information into unmistakable classifier information methodology. For wide range of utilization inconsistency identification, report characterization picture comment and substance particular for various information arrangements.

By observing the above talk, we discover the issue of single characteristic on dubious inconspicuous components sources and thought summary considering of the customer from record purposes of intrigue sources. As corporations within the real world want to recognize the Internet usage styles in their clients, these studies are beneficial to examine customer behavior and make strategies to enhance consumer delight. The Web usage mining for consumer conduct analysis has many real global utilities. Building a framework which could cater to the desires of organizations for user behavior analysis is a difficult mission. However, it's miles very useful to business community to make expert choices. The consumer conduct evaluation desires one-of-a-kind kinds of algorithms. Different strategies came into lifestyles so as to have user behavior analysis. However, we felt that a framework that can offer extensible functions to have user conduct analysis is wanted. In this paper, we proposed and applied a hybrid framework, i.e., extensible Web usage mining framework (XWUMF), which is familiar but gives placeholders for diverse destiny technologies. The framework is extensible and even supports personalized settings for user behavior analysis based on different patterns with respect to attributes. To increment multi-class name introduction with high-dimensional information continuously information applications, a superior framework is required to process diverse qualities. So in this paper, we recommend to make extensible and classification by pattern-based hierarchical clustering (ECPBHC) to describe record participates in light of properties in uncertain data streams with conceivable and ID formal parameters.

This paper presents extensible and classification by pattern-based hierarchical clustering (ECPBHC) method to improve unpredictable cross section to give widely low-level informational index portrayal. It is association-based way to deal with getting to immaterial information show in gathered information with various characteristics in view of likeness highlights. This investigation just partners the gap between the technique of information clustering and that of Web association examine. It moreover extends the ability to collection framework for particular information, which has not procured much thought in the creative works. Methodology of the pack gathering approach showed up in Fig. 1 with relative segments in aggregate social databases.

Despite the issue of collection particular information that is examined, thus, the suggested framework is by and large with the objective that it can furthermore be successfully used to other information types. The main goals of the suggested strategy give the following relational aspects:

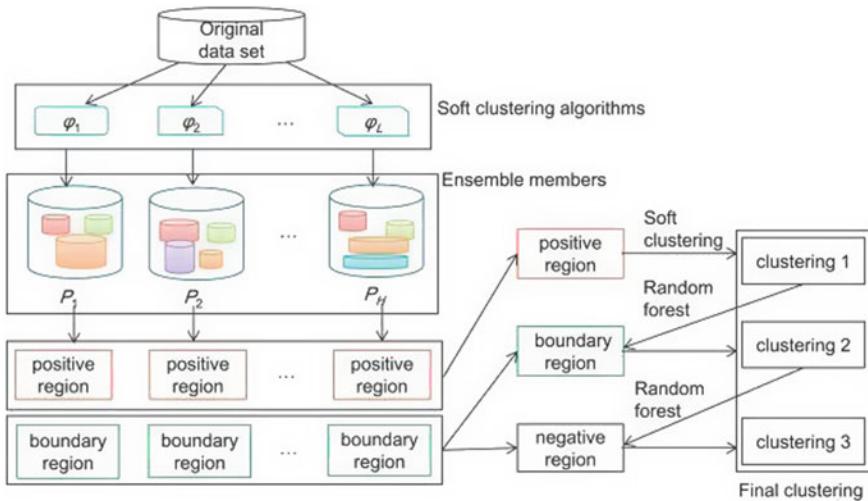


Fig. 1 Cluster data analysis for different data attributes

1. Briefly explain about clustering with regard to several features.
2. Respective process pertains to clustering with different features.
3. Application growth with regard to multiple features dividing.
4. Simulate the results acquired for different features with relations.

2 Review of Related Work

Vythoulkas and Koutsopoulos et al. employed neural networks, approximate reasoning, and fuzzy set theory for modeling choice behavior. They assumed that simple rules are used by decision makers. Rules and rule weights are used in the process of behavior analysis. Botha and Solms et al. used trend analysis and fuzzy logic for modeling behavior of different intruders. Anderson et al. focused on the analysis of user behavior with respect to fall in human life. Fuzzy logic and one-class person are used to detect fall behavior of humans using pre-recorded videos. Adadeh, Mohamadi, and Habibi et al. used genetic fuzzy systems for analyzing malicious users. They employed iterative rule learning to have know-how on the user behavior. A good survey of fuzzy Web mining can be found in various techniques pertaining to fuzzy Web structure mining, fuzzy Web content mining, and fuzzy Web usage mining. Velesquez et al. combined both Web usage mining and eye-tracking technologies for classifying Web site key objects.

Conti et al. studied user behavior pertaining Android application usage. They focused on user actions and the trends in the user behavior in using Android applications. Vu et al. focused on travel behavior of tourists. They used the notion of

geo-tagging photographs for user behavior analysis. Abello et al. made a survey of semantic Web technologies used for online analytical processing (OLAP) which can be used for user behavior analysis. We introduced and proposed a hybrid framework which is generic but provides placeholders for various future technologies. The framework is extensible and even supports personalized settings for user behavior analysis.

3 Background Approach

Sequential Web usage miner (SWUM) is used mining of Web usage patterns from different Web-oriented categorical datasets. The overview of general Web usage mining is shown in Fig. 2.

Web log information gathered from net servers is subjected to preprocessing after which user behavior analysis if you want to obtain enterprise intelligence. In this technique, the Web log statistics comes from distinctive sources. The information is preprocessed to enhance the statistics by handing missing values. Then, the data is subjected to usage mining, pattern discovery, and pattern analysis.

It generates patterns that reflect consumer conduct. First of all usage time of the Web pages is computed primarily based at the facts furnished within the dataset. The utilization time is considered to clear out processing through the use of MinTime parameter. The minimum self-assurance gives similarly statistical degree to have satisfactory styles. The seek space is pruned using MinTime and minimum self-belief parameters. This process can enhance the overall performance of set of rules for person behavior evaluation.

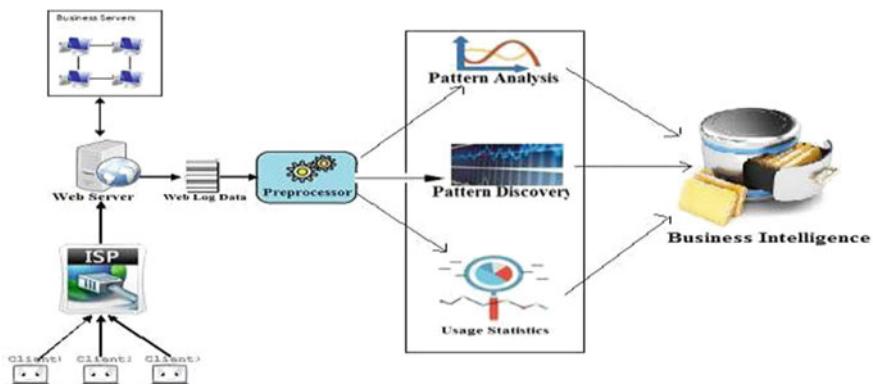


Fig. 2 Sequential Web usage framework for cluster data analysis

4 ECPBHC Method

In this section, we discuss about multi-attribute clustering specification, this approach actual listening with different attributes.

4.1 Basic Procedure for Data Summarization

Let $C = (c_1; c_2; \dots; c_N)$ be a combination of data relations with N details factors and $C = (1, 2, \dots, n) N_g$ be a team selection with M cluster analysis, each of which is referred to as an selection individual. Each platform clustering earnings a combined with categories $X_i = \{X_1^i, X_2^i, X_3^i, \dots, n\}$, such that $\cup_{j=1}^{k_i} C_j^i = C$, where k_i is different selection of cluster with different parameters. For each $\times 2 C$, $X(c)$ characterizes the combined brand similarity with factor c with cluster sequence.

In the i th clustering, $X(x) = "j"$ (or " X_j^i ") ifc X_j^i . This partition gives primary assets π^* of a complete set C , which contains grouped attributes with same attributes π [1, 2].

So, the basic cluster formation from different attribute clusters with suitable data with consensus learning functions based on the results with similar attributes procedure shown in Fig. 3.

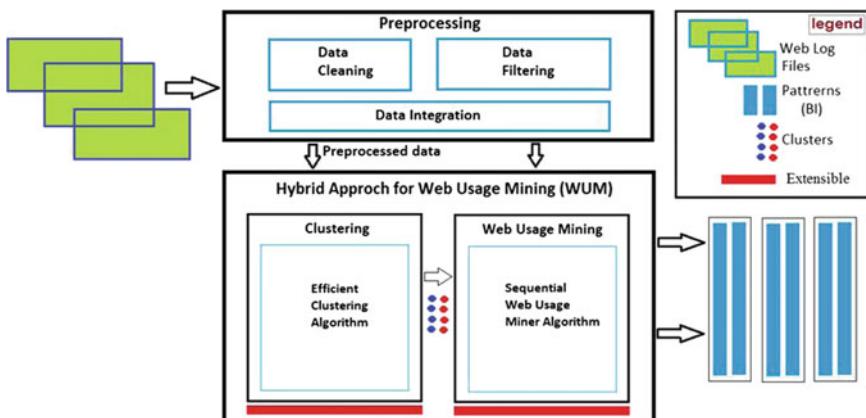


Fig. 3 Proposed approach procedure to access multi-attributes

4.2 Grouped Creation Approach

It is the essential idea to frame distinctive characteristics in mix with same relations. In bunching, singular qualities over extra information streams. Chosen characteristics accept numerous conditions with comparable highlights in view of customer prerequisites. In this circumstance, chose customers work the general framework change in light of bunch comes about. Significantly, a few characteristics prescribed the present traits in gathering approaches with the scope of specific progressive social qualities. At long last progressive highlights were utilized to portray specific gathering necessities with various multi goals.

4.3 Functions Identified with Consensus

Out of general traits, haphazardly select assembled highlights have been intended for accessible data with quality parcel. Utilizing Markov chain lattice development has comparable characteristics orchestrated in psychological capacities. A portion of the component-based methodologies with group examination changes working qualities continuously information streams for point-by-point classification. In Conesus, lattice arrangement with immediate and circuitous named developments.

4.4 Coordinate Technique

In coordinate approach, depending qualities are people for choosing social mark i.e. furthermore, number of properties in I with multi destinations in relations for various arrangements utilizing accord work. To give comparable qualities gathering for arbitrary determination from various informational collections. In Markova chain, demonstrate framework distinctive developments with qualities in light of Euclidian separation between all traits in information streams [3, 4].

4.5 Outlier Information Group for Properties

From the system of direct strategy with lattice development and characteristic game plan with comparable traits in relations. Anomaly development in view of qualities with numerous goals in various accords for gathering chosen includes in ongoing credits to distinguish exception from relations.

5 Experimental Setup

In this section, we formulate the performance analysis of proposed approach ECPHC with comparison of fuzzy cluster on real data streams. To develop this application, we use JDK and Net beans for user interface construction to upload data sets and perform single attribute classification and multi-attribute object classification from real-time data streams. Sample data sets are shown in Fig. 3.

Experimental Results: As shown in Fig. 3, proposed approach gives accuracy of real world entity in different data sets like accident, diabetes, economy ratings, student marks' performance in different formations. Sample results for after performing proposed approach for different data attributes are shown in Fig. 5.

As shown in Fig. 4, first upload data sets to our proposed approach for attribute selection with different relations, process each attribute as data point, and then perform Euclidian distances between attributes for efficient data classification with multi-attributes. Time efficiency for our proposed approach is shown in Fig. 5, different data sets like accident, diabetes, with multi-attributes in recent feature selection with randomly progress real-time data streams .

Furthermore, MAMG works constantly better than its competitors with all different selection measurements, while CO + SL appears to be the least effective. Realize that a bigger selection outcomes in an enhanced perfection, but with the trade-off of runtime.

Effectiveness comes about show up with ECPBHC with various multi-objects appeared in Fig. 6. To begin with, we make the restricted score to catch the territory vulnerability in light of each delineation near to data perform, and after that,

Fig. 4 Sample data sets with different attributes

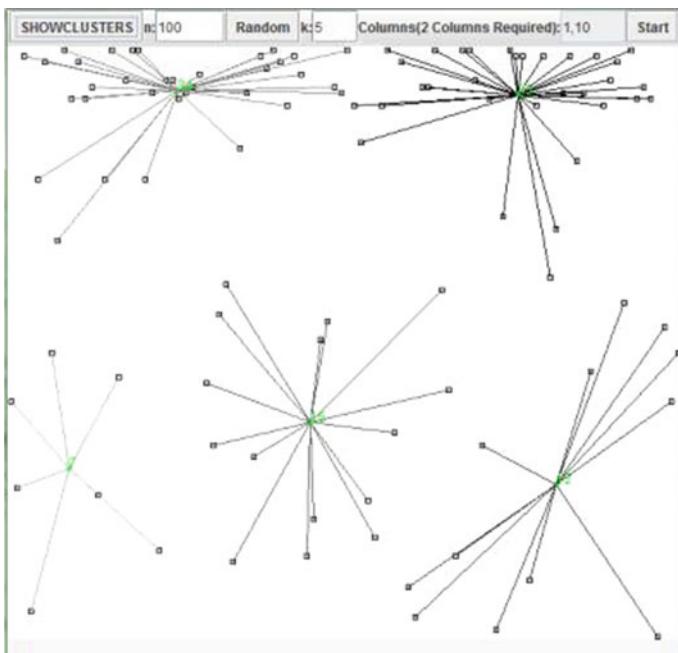


Fig. 5 Clustering with multi-attribute similarity measure for Web data

deliver a fetched one-class classifier by consolidating the vulnerability information into an ECPBHC with SVM-based examining system. Second, we make upgrade vector-based accumulation procedure to abridge the comprehension of the purchaser over the history pieces. Wide appraisals have uncovered that our unverifiable one classification contemplating can get a magnificent execution and are less touchy to complain regarding the standard one-class SVM. The appraisals furthermore show that the help of vector-based accumulation procedure can well lessen the comprehension of the purchaser regarding accentuate focused gathering method for idea rundown learning (Fig. 7).

6 Conclusion

In this paper, we propose and develop extensible and classification by pattern-based hierachal clustering (ECPBHC) for information discovery in different information resources with multiple object alignment in group relational information angles. This paper presents novel ECPBHC to classify information depending on different features from multiple perspective information resources. It constructs and converts matrix development into feature partition depending on chart procedure. Our trial outcomes give efficient and efficient approaches to set up information sets to measure features

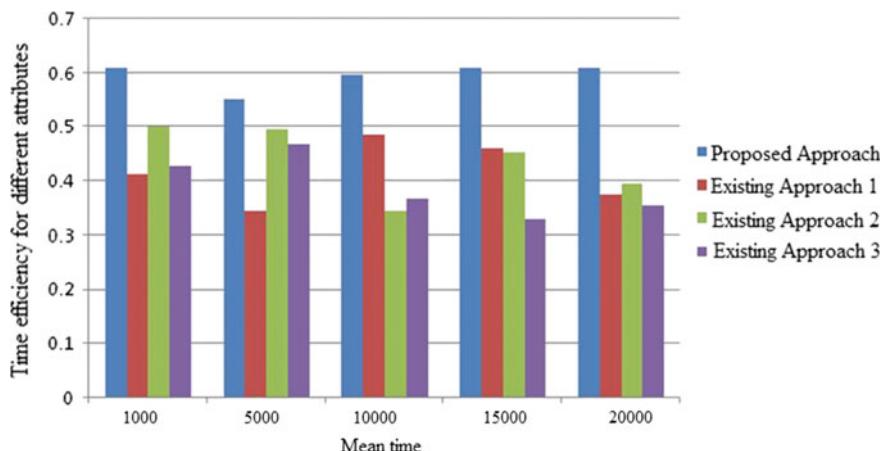


Fig. 6 Performance of time results with different attributes

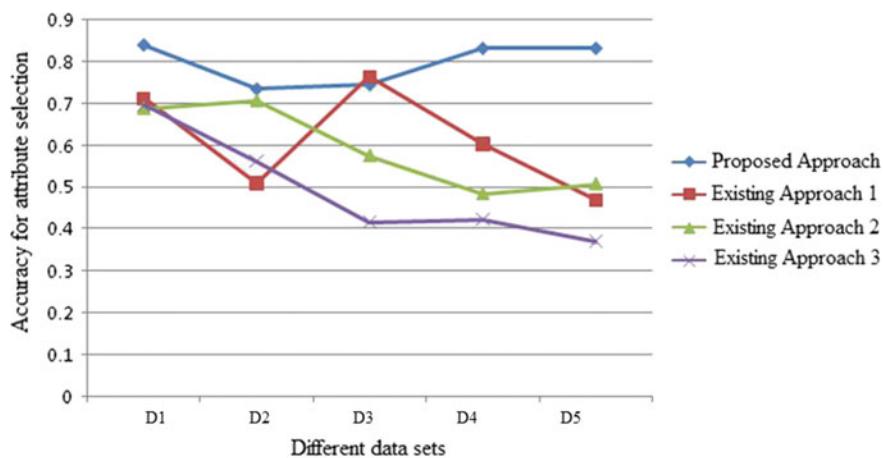


Fig. 7 Accuracy to access data from different data sets

and combine those features using link-based technique. It gives efficient leads to multiple feature combination from group relational databases with semantic information structure with likeness measures with feature partition. As further improvement of our suggested approach is to apply ECPBHC in different types of real-time information sources using advanced machine learning methods. We also design best assessment method to extract information depending on multiple attributes.

References

1. Suneetha, K.R., Krishnamoorthi, R.: Identifying user behavior by analyzing web server access log file. *IJCSNS* **9**(4), 1–6 (2009)
2. Chen, L., Wang, C.: Continuous subgraph pattern search over certain and uncertain graph streams. *IEEE Trans. Knowl. Data Eng.* **22**(8), 1093–1109 (2010)
3. Geng, B., Yang, L., Xu, C., Hua, X.: Ranking model adaptation for domain-specific search. *IEEE Trans. Knowl. Data Eng.* **24**(4), 745–758 (2012)
4. Hido, S., Tsuboi, Y., Kashima, H., Sugiyama, M., Kanamori, T.: Statistical outlier detection using direct density ratio estimation. *Knowl. Inf. Syst.* **26**(2), 309–336 (2011)
5. Agichtein, E., Brill, E., Dumais, S.: Improving Web Search Ranking by Incorporating User Behavior Information. ACM, pp. 1–8 (2006)
6. Stummea, S., Hotho, A., Berendt, B.: Semantic Web Mining State of the Art and Future Directions, pp. 1–20. Elsevier, Amsterdam (2006)
7. Zaiane, O.R., Luo, J.: Towards evaluating learners' behaviour in a web-based distance learning environment. *Knowl. Data Eng.* **26**(2), 1–5 (2001)
8. Bovoloa, F., Camps-Vallsb, G., Bruzzonea, L.: A support vector domain method for change detection in multitemporal images. *Patt. Recogn. Lett.* **31**(10), 1148–1154 (2010)
9. Lian, X., Chen, L.: Similarity join processing on uncertain data streams. *IEEE Trans. Knowl. Data Eng.* **23**(11), 1718–1734 (2011)
10. Huffel, S.V., Vandewalle, J.: *The Total Least Squares Problem: Computational Aspects and Analysis*. SIAM Press (1991)
11. Gunn, S.R., Yang, J.: Exploiting uncertain data in support vector classification. In: Proceedings of 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, pp. 148–155 (2007)
12. Jiang, B., Zhang, M., Zhang, X.: OSCAR: One-class SVM for accurate recognition of CIS-Elements. *Bioinformatics* **23**(21), 2823–2828 (2007)
13. Jin, R., Liu, L., Aggarwal, C.C.: Discovering highly reliable subgraphs in uncertain graphs. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 992–1000 (2011)
14. Sai, S., Reddy, S., Mamatha, Ch., Chatterjee, P., Nagarjuana Reddy, S.: Contemporary supply chain and inventory data management using data analytics. *Int. J. Mech. Eng. Technol. (IJMET)* **8**(12), 290–294 (2017)

IAAS Service in the Public Domain: Impact of Various Security Components on Trust



Archana B. Saxena and Meenu Dave

Abstract Cloud computing a glorified paradigm that has seen a sharp rise in the last few decades and comes up as the fastest growing IT segment. This computing technology has stretched the prevailing capabilities of computing resources by offering them on a sharing basis. Shared resources have made it affordable for medium and small-scale business to use this latest technology and upgrade themselves in term of technology at par with big business houses. Through this technology, an individual or group can use the latest IT tools/technologies/platforms without any huge investment, licensing and installation glitches, and training and administration concerns. As more and more individuals and societies are joining this stream, the question has become more relevant is the cloud safe? The biggest promises made by the technology become counterfeit when data leakage incidences surfaced. By keeping in mind the utility of the technology, more and more researchers are trying to address the concerns faced by the technology. Important challenges for the technology are security, privacy, and trust. These issues are interrelated. The present study is an inspection to find the impact of various security components on trust. To achieve the stated objective, a study of concerned literature was conducted followed by the primary survey between free and paid cloud consumers through an online questionnaire. Collected responses are analyzed through SPSS to find out the impact of each component on trust.

Keywords Cloud computing · Security · Trust · Infrastructure as a service · IaaS

A. B. Saxena (✉)
Jagan Institute of Management Studies, Delhi, Delhi, India
e-mail: archanabsaxena@gmail.com

M. Dave
Jagannath University, Chakshu, Jaipur, Rajasthan, India
e-mail: meenu.s.dave@gmail.com

1 Introduction

A relatively new technology “cloud computing” is using old computing techniques like grid computing, utility computing, timesharing, SOA architecture, and virtualization [1]. The success rate of this technology is because of its architecture that allows sharing of resources through the internet from different geographical locations. This unique feature of the technology made it possible for a middle and small business house to have almost at par technological advancements to their big business counterparts. The best part is that, without any hardware investment, administrative burden, installations glitches and maintenance efforts. The best part of this computing paradigm is that financials are totally based on usage and scalability of resources is at the ease or demand of consumers. The technology is capable of providing hardware resources, load balancers, storage space, and networks through infrastructure as a service (IaaS), operating systems and application development framework can found through platform as a service (PaaS) while software as a service (SaaS) allows users to use applications without installing them on their own machine [2]. These applications are directly served to the consumers from data providers’ location called “data centers.” Data center’s pool large amount of high configuration resources that are executed, managed and administered by the cloud provider [3]. These centers are the key points to deliver Web-based cloud services to consumers’ workstations. At consumers’ workstation, these services can be accessed through the Internet via variety of devices like laptop, desktop, PDA, phone, etc. [3]. The architecture of the technology has various characteristics that leave keyholes for security like multi-tenancy, elasticity, multiple stakeholders, third party control [4] and most important threat is its delivery mechanism Internet. Due to these security punctures, this fastest growing technology is threatened with various challenges that are hampering the growing trend and impeding users from adopting it. The major challenges faced by the technology are security, privacy, and trust [5, 3]. It has been observed during the review that security and trust are loosely related to one another in cloud concerns [6]. Few trust models are also found in the literature that links trust with security and generates trustworthiness of the provider based on a few components that are part of security assurance. This model generates a trust value by keeping the following measures like authentication, authorization, identity management, data protection, and virtualization [7]. A model is proposed that used to enhance security and interoperability [8]. Trust model that is based on a few security components like customer and service integrator that uses security as a base component to test and enhance trust [9]. Trust model that is based on the mobile agent to ensure security measures and monitoring activities [10]. The literature is also flooded trust models that use security as one the component to evaluate or enhance trusts like SLA, compliance, portability interoperability geographical location, customer support, performance, federated IdM, security, user feedback, service delivery, and deployment models [11]. There are trust models that are not directly using security terms but components that they are using are indirectly related with the security like trust model based on QoS uses

components like availability, reliability, turnaround efficiency, and data integrity. Data integrity has an angle that linked with the security of data [12].

2 Research Gap

Even though a lot has been written about security and trust in cloud environments, plenty of security-based frameworks to address trust are available in the literature, still following gaps have been observed by the authors:

- Presently, available trust factors are using security as a component in the computation of trust without realizing the impact of security on trust.
- Security has many more subcomponents, what is the impact of these subcomponents on trust.
- Expect security there are many more components that have an impact on trust. There is no evidence of such a framework in the literature that uses these different components along with security in trust computation.
- The awareness level of consumers about various security assurance components like standards, certifications, and guidelines.
- The relevance of standards, certifications, and guidelines for different cloud consumers.

3 Objective

This paper is an extension and improvement of existing work that authors are doing in the area of “security and trust.” The main objective of this research is to determine the impact of various security components (security, governance, SLA, audit, and diverse) on trust [13].

Security is not a simple term that can define easily. It is rivulet that can be defined with a set of components. The selection of security components for this research will be based on their impact on trust. Different security components that play an important role in building or slashing trust are security, governance, SLA, audit, and diverse [13].

The selection of components for the evaluation of impact will be based on the following parameters:

- Whitepaper by CSCC (Cloud Standard Customer Council) for cloud consumers regarding security requirements of public clouds [14].
- Guidelines for SLA between consumer and provider by CSIG (Cloud select Industry Group).
- Literature related to security essentials for IAAS services in public cloud.

In addition to the central research questions, there are few more questions that can be reveal through this paper:

- Are cloud consumers aware of standards, certifications, and guidelines issued by legal bodies in reference to cloud computing?
- Do people give preference to provider that attains guided standards and certifications?
- A provider that follows the guidelines of private bodies is preferred among others or not.

4 Research Methodology

To find a solution to research problem, a primary survey is conducted among free and paid cloud consumers. An online questionnaire is prepared through Google forms. The objective of the questionnaire's framed questions is to capture the respondent's opinion about trust and factors that can impact consumers trust in a cloud provider. The questionnaire is distributed among free and paid cloud consumers and responses are recorded online. After cleaning and filtration, online responses are migrated in SPSS and analyzed.

The framing of questions is done after deep analysis of trust, affected components, models and related literature, discussion with SME and authors own knowledge about the area. Reference has also taken from authors previous work about "Factors affecting trust" [Dr. Meenu Dawe and Archana B. Saxena].

5 Findings

A mix of responders (Table 1) from large (48.9%), medium (22.2%), and small (28.9%) has taken the initiative and given the response to the questions formed.

Most of the responders are using either public (62%) or hybrid (11.1%) deployment model of (Fig. 1) cloud computing. Which is good for our study because our study is based on IAAS model for public model. A variety of users using different cloud services like 57% are using IaaS services, 44% are using SaaS, and 11% are using PaaS services have given their reviews on the impact of various security com-

Table 1 Distribution of respondents on the basis of their company size

		Percent	Cumulative percent
Valid	Large (>500)	48.9	48.9
	Medium (100–500)	22.2	71.1
	small (1–100)	28.9	100.0
	Total	100.0	

Source Primary survey

Fig. 1 Distribution of respondents on the basis of cloud computing models

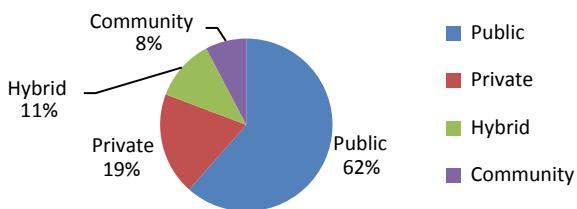


Fig. 2 Awareness about standards and certifications among respondents

Awareness_Standards_Certifications_Guidelines

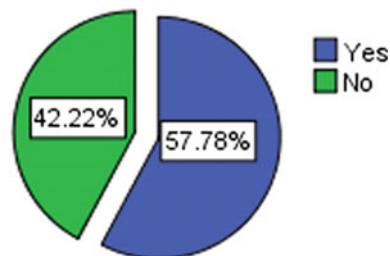
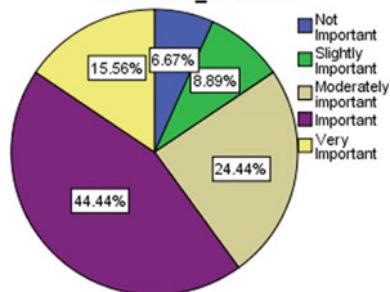


Fig. 3 Importance of defined Standards for the cloud consumers

Relevance_Standards



ponents on trust. More than half of responded population is using IAAS services, so analysis would more accurate.

Among the respondent population, 58% is aware of standards, certifications, and guidelines in reference to cloud services. Forty-two percentage are not aware of these norms issued for cloud services.

Despite awareness, a variation is found in the relevance of standards, certifications, and guidelines in trust. There is 78% (Fig. 2) population from data sets, considered it essential for a provider to follow standards defined by regulating authority or legal system where cloud services are in use. Whereas only 44.44% considers it significant to attain standards related to cloud (Fig. 3). Whereas only 33.33% finds a need to follow the guidelines by the cloud provider (Figs. 4 and 5).

The objective of this questionnaire is to analyze the impact of various components like security, governance, audit, SLA, and diverse on trust. Each component

Fig. 4 Importance of attaining certifications for the cloud consumers

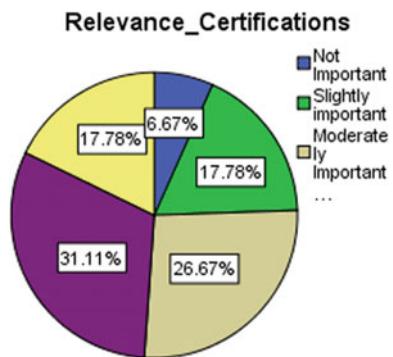
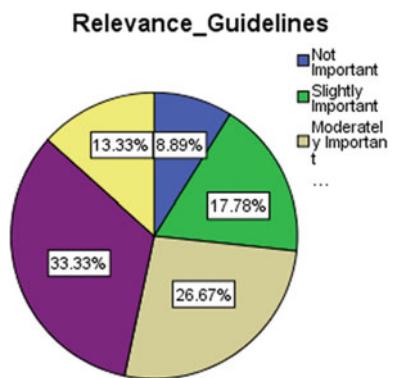


Fig. 5 Importance of guidelines by the cloud provider



has certain subcomponents; in fact, questionnaire includes questions that check the impact of each subcomponent on trust. From Fig. 6, it can be analyzed the most of the respondents have given the maximum weightage to the security which comprises following components: (data security, privacy policy, authentication and authorization, network and Internet security, cloud application security and security of physical infrastructure) in trust formation for cloud services. There could be plenty of reasons that are taking rounds in mind of respondents in selecting security as prime components that impact trust. They are:

- Nature of their information and data, that they saving on cloud resources.
- Data leakage incidents, becoming a regular feature of in newspaper headlines.
- There past experiences where they had financial or reputation loss because of security lapse.

Respondents have given almost similar weightage to the next two components governance (Mean: 3.7333) and SLA (3.7067). Governance has given only 266 more preference than SLA, which is because SLA is treated as contractual document while governance has more impact on trust because it includes subcomponents like risk policy, risk compliance, and exit process. For consumers, they are more relevant than few performance, security, and data management metrics. Although responsibilities

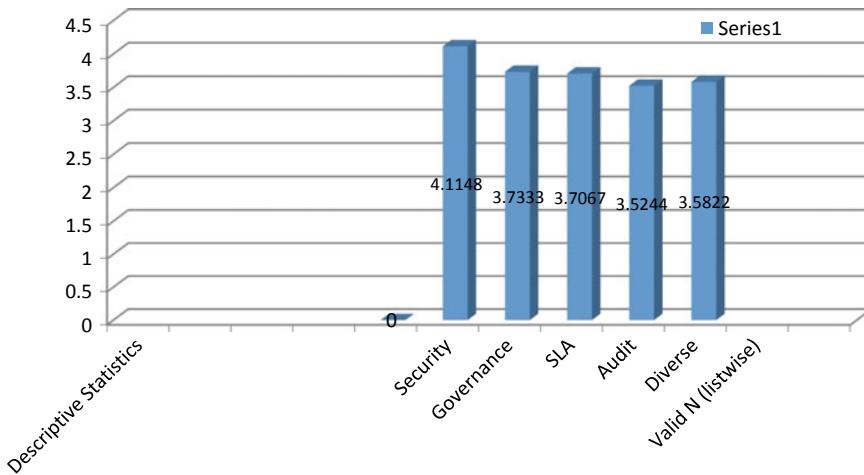


Fig. 6 Relevance of components having as having an impact on trust

and penalties are also listed in SLA, maybe respondents believe in precautions then curable measures.

Among the remaining two components audit (Mean: 3.5244) and diverse (Mean: 3.5822). Although the difference is very minute, only 578 respondents have given more concern to diverse which includes subcomponents like technology stack, capacity, support, and help desk, cost-benefit equation, and current customers' feedback. This is because cloud consumers have given more importance to feedback, capacity, and technology stack as compared to internal, external audits, and log details.

Figure 7 can very well explain the relevance of each component by respondents has different deployment models.

6 Conclusion and Future Scope

This research paper is trying to find the impact of various components and subcomponents of trust. Consumers' awareness and importance of standards, certifications, and guidelines among consumers are also collected and analyzed for further. The findings of this research paper will be used to develop an evaluative trust model that will be based on security coverage offered by the cloud provider. On the basis of their impact on trust, components will be arranged in the trust model for calculation of trustworthiness of cloud provider.

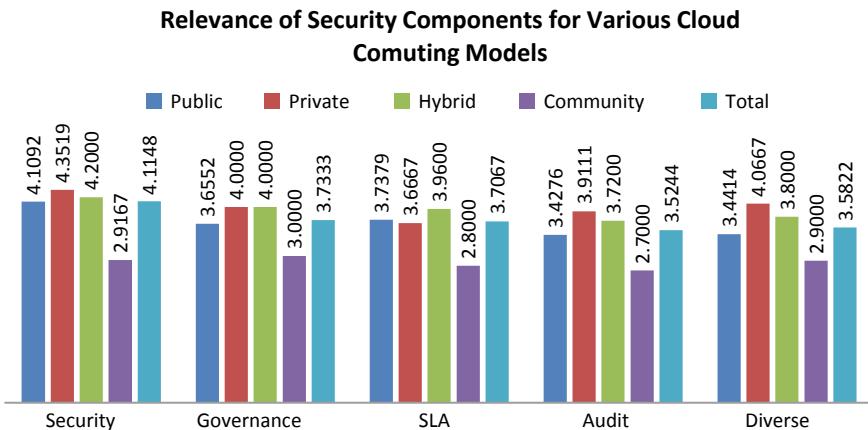


Fig. 7 Mean value of subcomponents calculated for each component based on relevance marked by respondents

References

1. Lee-Post, A., Pakath, R.: Cloud computing: a comprehensive introduction. In: Srinivasan, S. (ed.) *Security, Trust and Regulatory Aspects of Cloud Computing in business Environment*, pp. 1–23. IGI Global, Texas (2014)
2. Jadeja, Y., Modi, K.: Cloud computing—concepts, architecture, and challenges. In: ICCEET—International Conference on Computing, Electronics and Electrical Technologies, pp. 877–880. IEEE, Kumaracoil, India (2012)
3. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. *J. Internet Serv. Appl.* 7–18 (2010)
4. Tianfield, H.: Security issues in cloud computing. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1082–1089. IEEE, Seoul, Korea (2012)
5. Pearson, S.: Privacy, security, and trust in cloud computing. In: Pearson, S., Yee, G. (eds.) *Privacy and security for cloud computing*, pp. 3–42. Springer, London (2013)
6. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. *J. Network Comput. Appl.* 1–11 (2011)
7. Jagtap, A., Kamanache, A., Sharma, Y., Rathi, S.: Trust model for cloud computing. *Vishwakarma J. Eng. Res.* 251–254 (2017)
8. Li, W., Ping, L.: Trust model to enhance security and interoperability of cloud environment. In: First International Conference on Cloud Computing 2009, pp. 69–79. Springer Nature, Beijing (2009)
9. Takabi, H., Joshi, J.B., Ahn, G.: Secure cloud: towards a comprehensive security framework for cloud computing environments. In: 34th Annual IEEE Computer Software and Applications Conference Workshops, pp. 393–398. IEEE Digital eXplorer (2010)
10. Hada, P.S., Singh, R., Meghwali, M.M.: Security agents: a mobile agent based trust model for cloud computing. *Int. J. Comput. Appl.* 12–15 (2011)
11. Habib, S.M., Hauke, S., Ries, S., Muhlhauser, M.: Trust as a facilitator in cloud computing: a survey. *J. Cloud Comput. Adv. Syst. Appl.* 1–18 (2012)
12. Manuel, P.: A trust model of cloud computing based on quality of services. *Ann. Oper. Res.* 1–12 (2013)

13. Dawe, D., Saxena, B.A.: Loss of trust at IAAS: causing factor and mitigation techniques. In: International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN 2017) (pp. 1–8). IEEE, Gurgaon, Haryana (2017)
14. Baudooin, C., et.al.: Cloud Security Standard: What to Expect and What to Negotiate Version 2.0. Cloud Standards Customer Council (2016)

Serving the Dermatologists: Skin Diseases Detection



Savy Gulati and Rosepreet Kaur Bhogal

Abstract Skin diseases are one of the most common diseases found among the humans that have adverse effect on them, whether it is a small bump, which might lead to bothersome or a vicious structure and then to mortality. Despite of this, these diseases are generally taken leniently by majority of the strata. Skin diseases are a huge burden on the world, and there is an alarming need to get it into control at early stages, so that effective measures can be taken up as early as possible. Moreover, the number of dermatologists is quite low as compared to the subjects, which is a serious issue. So, computer-aided diagnosis is a boon in such scenarios as they not only dilute the job of dermatologists but also add effectiveness to their work by reducing the diagnosing time. Further, it can make easy for experts to suspect those lesions which are difficult for naked eyes. These CAD systems can also act as a standalone system to diagnose diseases at early stages. It can be fruitful for the people of rural areas which are deprived of easy access to dermatologists. In this research, skin disease detection based on electrical properties and image processing has been considered. Apart from this, comprehensive review on numerous techniques to detect skin diseases such as skin cancer and other skin diseases (acne, psoriasis, orofacial diseases, warts, erythematous-squamous diseases, tinea corporis, vitiligo, eczema, legs, and feet ulcers) is carried out.

Keywords Skin diseases detection · Electrical properties · Image processing · Skin cancer · Other skin diseases

S. Gulati · R. K. Bhogal
Lovely Professional University, Phagwara, India
e-mail: savygulati99@gmail.com

R. K. Bhogal
e-mail: rosepreetkaur12@gmail.com

1 Introduction

Human skin is the soft outermost layer of the human body. It contributes to the major area as it is the largest organ spread throughout the body, putting up 16% of the body mass [1]. Skin acts as a sheath which protects human body from microbes, maintains body temperature, allows the sensation, and prevents unbalanced losses of essential solutes and liquids such as water [2]. There are three essential elements of the human skin, namely hypodermis, dermis, and epidermis. Firstly, hypodermis is the innermost layer mainly for fat storage. Secondly, dermis exists between the middle of the hypodermis and the epidermis layer, and this layer mainly provides elasticity to the skin. Thirdly, epidermis is the outermost layer, whose first job is to serve protection; also, it is waterproof and has a regenerative capability from the wounds [3]. Skin disease is an abnormal condition of a skin, ranging from a tiny red bump to large evil structure. There are many reasons which contribute to skin disorders, such as infections, genetical issues, stress, environmental, allergies, and many more [4].

It is evident from the well-known study carried out at global level [5, 6] that burden of skin disorders ranked fourth among the top non-lethal diseases around the world. It is also observed that though the mortality rates are less for skin diseases, their botheration to subjects is quite stressful. If particularly consider the case for USA [7], every one person in four suffered from a skin disease; in addition to this, it was also noticed that the money spent on skin diseases was comparable or more than that of the chronic diseases. Similarly, if considered for India, [8] investigates that 60% of Indians are within the trap of skin diseases, also different age groups suffer from various diseases. Thus, today skin diseases and lesions are most common and require an effective way to deal with. There are both invasive and noninvasive conventional methods to investigate skin diseases. Biopsy is an invasive method, in which skin piece is removed for diagnosis purpose, whereas dermoscopy is a non-invasive method which provides the clear and zoomed image of the affected portion. Computer-aided diagnosis incorporates dermoscopic images with the specific PC-held framework, and this optimizes and speeds up the task of dermatologists, thus act as a right hand for experts. Apart from this, CAD systems can also act as standalone systems which can intimate subjects about their skin diseases at early stage, thus also benefiting the people of rural areas which are deprived of quality facilities.

In this research, comprehensive review on several techniques proposed by various researchers to detect skin diseases has been carried out which include skin cancer, and some other skin diseases such as acne, psoriasis, orofacial diseases, erythematous squamous diseases, warts, tinea corporis, vitiligo, eczema, legs, and feet ulcers. Section 2 is devoted to electrical properties-based and image-based approaches to detect skin diseases. Section 3 includes an overview of various skin diseases, and review of the techniques used for detection of these skin diseases has been performed. Section 4 includes conclusion and future scope.

2 Approaches to Detect Skin Diseases

Here concentrating on two approaches to detect skin diseases; first is by utilizing electrical properties, and second is by using image processing.

2.1 Use of Electrical Properties

Biological electrical properties measurement is an umbrella term, many studies that have been conducted are evident that these can be employed to differentiate between sound and diseased skin. P. B. Manoorkar, D. K. Kamat, P. M. Patil proposed a prototype which has AD5933 Network analyzer for providing the electrical impedance, and it is based on microcontroller. Here, impedance measurement concept has been exploited that the impedance of both the diseased and healthy skin is different when both tested at same frequency, and this is due to the fact that rashes, irregular structures, bumps, allergic reactions, redness, irritability, and all such issues lead to change in impedance as diseased skin flourished by local blood supply; moreover, nucleus of cancer cells is bigger than the normal skin's intra-cellular spaces, which assist to perform distinction between normal and diseased skin. Analysis of impedance is made using four parameters: magnitude, phase, real part index, and imaginary part index. This approach can detect both cancerous (basal cell carcinoma, squamous cell carcinoma, malignant melanoma) and non-cancerous diseases such as acne, psoriasis, sickle cell anemia, scabies, rubella, mouth diseases, and leprosy. It has been evident that healthy skin has more magnitude as compared to the diseased skin. This technique is effective as it consumes less power, money, and is quite handy. It provides a detection rate of 75%. No doubt that the system is well capable of recognizing diseased skin, but the provision of disease classification is not provided [9].

Also, Zogkas et al. [10] employed the same concept of DermSense that is electrical impedance spectroscopy and proposed the prototype which proves to be fruitful for the identification of malignant melanoma. Its features include portability, easy functionality, and non-bulkiness. Prototype consists of data electrode arrangement to measure impedance from the skin, AC signal generator to provide electric current to the skin, and lastly a data acquisition unit. It is evident that samples collected from the cancer-affected skin show larger variation as compared to that of the samples of normal skin, also the cancer samples are not symmetric; thus, the impedance of cancer skin diverges more than that of the healthy skin. Apart from this, it is observed that the standard deviation of the former is 7 times more than that of the latter. Thus, standard deviation and impedance act as performance measures for the melanoma detection. As this prototype was made for initial testing purpose, so it involves less test measurements. In spite of the productive results, this system could be optimized further by incorporating detection of other skin diseases as well. Mirbeik-Sabzevari et al. [11] use millimeter waves for the detection of skin diseases, for this experiment performed on healthy and diseased tissues of one hundred and one

subjects who have undergone surgeries at medical center of Hackensack University. Here the difference between the dielectric properties of normal and cancer affected person is considered at the millimeter range of frequencies. Components involved are coaxial cable, Hoboken analyzer. Framework involves measurement of reflection coefficients with the help of the coaxial cable; then, these coefficients are taken to the aperture of interest, followed by the measurement of desired permittivity values; then, Cole–Cole model is utilized to fit into these permittivity values, and the values which are not able to fit into these are rejected, and then, finally dielectric properties are examined using MANOVA. Apart from diagnosing normal and cancerous skin, samples from different patients also compared among themselves. Results include effective proofs in terms of change in dielectric properties of basal and squamous cancer tissues, also of normal skin tissues thus contributes to the detection problem. Other than this, it has been observed that not only water imbalance lead to dielectric changes or tumors but also changes in the biological structures and concentrations contribute to it. Further, this efficient approach can be extended to other diseases as well, which can promote optimization.

2.2 *Use of Image Processing*

Medical image processing is another noninvasive method whose popularity and efficiency is increasing at a faster pace. In this approach, image of the affected area is taken as an input and the diagnosed results are outputted by the system itself by processing that input image. This methodology is comfortable and easy to operate. However, whole of the task revolves around the image thus the prime need for such kind of systems is clear and a quality image so that the analysis can be correctly carried out. Basic steps included in image processing are preprocessing, segmentation, features extraction, features selection, and classification.

Preprocessing: In this stage, various operations are applied to remove artifacts present in the skin diseased image such as hairs, ink markings, air bubbles, noise shapes, ebony frames, uneven distribution, and reflection artifacts. Ultimately this stage will make the image ready for further processing stages and remove all the potential noisy factors.

Image Segmentation: Here image is partitioned into the group of pixels which are homogeneous with respect to certain criterion. Segmentation reveals the boundaries of the skin lesions, which are used to get the region of interest. Skin lesions are quite irregular, thus to segment them is a challenging task. This is the very crucial step as the wrong segmentation can deviate the classification results from the actual ones.

Feature Extraction: Diseased image's information is reduced to the components arrangement that is robust toward the lightning, lack of clarity, and camera positioning.

Feature Selection: At this stage, effective subset is chosen, by picking only those particular features which will result in fruitful classification of the skin diseases.

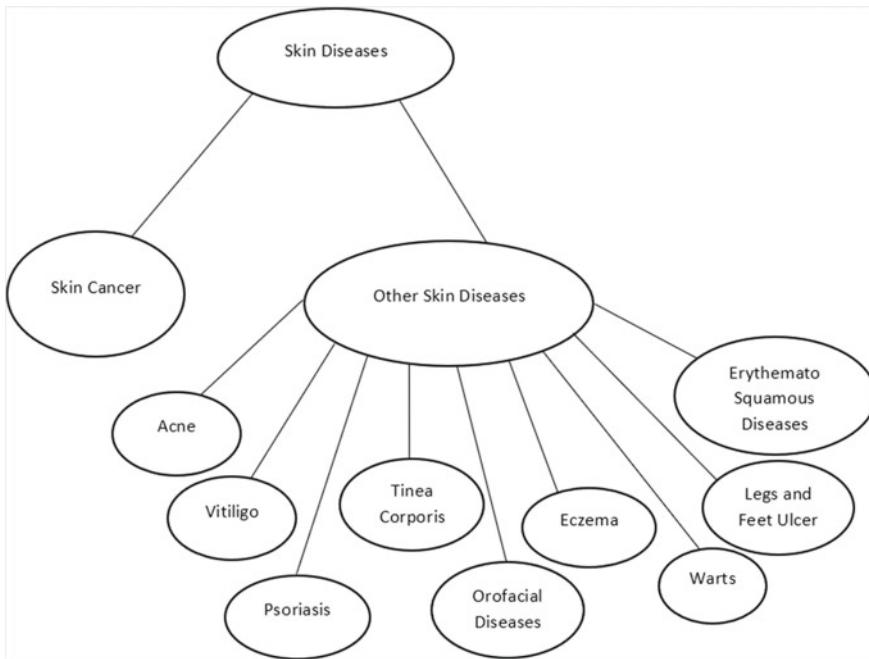


Fig. 1 Various skin diseases

Hence, this step reduces the redundancy, thus making the process of classification more robust and less time-consuming.

Classification: This is the final stage which involves the formation of classes of skin diseases on the basis of the selected features.

3 Skin Diseases

Various skin diseases which are considered here are stated in Fig. 1.

3.1 Skin Cancer

Skin cancer is quite a frequent cancer type, which is hiking at a faster pace. Majorly skin cancer is classified as melanoma and non-melanoma, whereas non-melanoma is further divided as basal cell and squamous cell carcinoma. Major reason which contributes to skin cancer is UV radiations. Malignant melanoma is the skin tumor which arises from melanocytes, whereas basal cell and squamous cell carcinoma arise

from Keratinocyte lineage cells. BCC is the least dangerous among these three, and with right treatment, this can be easily cured. On the other hand, SCC is the second most common skin cancer and may turn into a huge mass if not treated timely. SCC is severe but not as that of melanoma. Melanoma is the deadliest of all which is life-threatening [12]. Early detection of both melanoma and non-melanoma proves to be fruitful [13].

Chakraborty et al. [14] classified malignant and benign lesions by utilizing meta-heuristic supported ANN. Achieved results compared with two other techniques NN-CS and NN-PSO and observed that proposed NN-NSGA-II methodology outperforms among them. Suggested method can be optimized by including preprocessing and segmentation stages. In [15], Alquran et al. used SVM to classify the malignant and benign lesions. Results are obtained with all 11 features compared to those of the five PCA selected features, and it is evident that it later reduced the computation complexity. However, no mechanism for hair removal is given in this approach. Thus, addition of hair removal algorithm can make it more robust classification technique. Chakraborty et al. [16] in their work exploited NN-NSGA for classification but the features are extracted using discrete wavelet transform and then selection made by using PCA. Other than this, three diseases are classified, namely simplex lentigo, skin angioma, and basal cell carcinoma and the results compared with NN-GA and NN-PSO. It is observed that the proposed algorithm provides an edge over the other two methods. Given technique can further be refined by including both preprocessing and segmentation stages. Joseph and Paniker [17] proposed an efficient method for detection and removal of hair from images based on 2D-Derivative of Gaussian and Fast marching In-painting algorithm. This method involves two classifiers: first classifier (C1) checks lesion for benign and abnormal, if the lesion is found to be abnormal, then second classifier (C2) will check for the atypical or melanoma. This approach can be extended to other skin diseases as well. In [18], UIAin et al. exploited genetic programming to classify melanoma and non-melanoma. In this system, classification is performed by considering three scenarios, firstly by using only LBP features, secondly using only dermoscopic features, and lastly using both LBP and dermoscopic features. Classification results are best for the scenario utilizing both the features. Apart from this, GP classifier is compared to other six classifiers. It is observed that GP provides improved or equal results to other classifiers. Among dermoscopic features, it is also seen that blue-whitish veil and dark brown color features prove most efficient for the diagnosing process. The performance of this system can be hiked by including segmentation technique as non-interest regions can be discarded.

Waheed et al. [19] carried out classification by firstly using texture features alone, then color features alone, and lastly using both. It is evident that the results are better in last case. Performance compared with other four methods and seen that proposed method is best among them. System's accuracy can be improved by including dermoscopic features. Reshma and Priestly [20] identify melanoma stage using ABCD parameters and melanoma type using location of melanoma and correlation of other features. This technique can be more generalized by including diseases other than cancer. In [21], Jain and Jagtap proposed a method to classify healthy and diseased

skin. In this, geometric features are extracted to classify the skin. This work can be improved by detecting even the type of cancer. Udrea and Mitra [22] proposed GAN for the segmentation of skin lesion from the background as skin lesions are quite irregular in shape so their automatic segmentation is a quite challenging task. Generative network contains two different networks, namely generator and discriminator. Generator tries to produce an image similar to that of the input image while the task of the discriminator is to misguide the generator. And at one time, image is so proximate to the input image that discriminator will be failed to distinguish it. Generator used here is basically a Unet. Two approaches used for training: with rotation and without rotation, observed from the results that former outperformed the latter. Problem seen in this system is that it is difficult to segment small lesions, and if test images are converted to gray to apply Otsu segmentation to eradicate this problem, then it fails to identify multicolored lesions. Despite effective results, the framework is only for achieving the segmentation, but the ultimate goal is to classify the diseases. In [23], Agarwal et al. proposed the region growing method to segment the skin lesions. Two binary images are generated. Then, postprocessing is done to select the best from two. The one with higher solidity and extent is selected, and the other one is discarded. Segmented image is compared with the ground truth values and thus observed highest co-relation and overlapping of 97.93 and 91.13%. In addition to this, it has been concluded that computation time directly depends on the lesion size. Here also the classification of cancer has not been performed.

Takruri and Abubakar [24] proposed fusion of three SVM classifiers with radial basis function. Performance of proposed classifier is compared with the majority voting block and probability averaging fusion block. First SVM uses 21 GLCM features, second uses 52 curvelet features, and third uses 179 DWT. Accuracy and confidence distribution are calculated to measure the performance and observed that the proposed algorithm outperforms individual classifiers in terms of former, whereas comparable in terms of latter. Diseases other than skin cancer can be included which will make the system more diverse and fruitful. Ahn et al. [25] established RSSLS method which removes hairs from the image; then, using multiscale background detection is done utilizing super-pixel algorithm, boundary connectivity method for extreme cases. After this, image is reconstructed using sparse reconstruction followed by content-based error propagation. Lastly saliency based segmentation using Bayesian network is employed. Results of proposed segmentation are compared with six unsupervised lesion segmentation and seven unsupervised saliency detection methods. It is observed that suggested system shows an edge over them, but it fails to detect small lesions. Saliency information can be exploited to classify the lesion. In [26], Patel et al. used K -means clustering in combination with entropy to perform efficient skin lesion segmentation. Results measured using four parameters and thus get average RE, KI, JI, and CDR of 8.06, 94.83, 90.25, and 93.76, respectively. The system can be employed for classification of skin disorders. Munia et al. [27] segment skin lesion using combination of k-mean clustering and Otsu thresholding. After this, postprocessing is done using morphological operations and guided filter. Proposed method carries out classification using three classifiers out of which SVM with linear kernel performs the best having accuracy of 89.07%. Despite efficient results,

there is a lack of hair and bubble removal algorithm, which can further optimize the system's performance. Fonseca-Pinto and Machado [28] used LBP histograms after performing BEM decomposition. Then, optimized features are selected using AdaBoost and given to weak classifier. The method is implemented using both with and without postprocessing and compared to five other algorithms. It can be seen that the proposed method shows satisfactory results even without preprocessing. This algorithm can be extended to mobile-based application.

Serkan et al. [29] created prototype using Raspberry Pi and achieved more than 90% specificity and sensitivity for all three categories: common nevi, benign nevi, and melanoma. Algorithm can be made more generalized by incorporating more number of skin diseases. Firmansyah et al. [30] calculated dermoscopic features of the pictures taken using mobile phone, and based on the TDS score, lesion is categorized. All the processing is done within the smartphone. This algorithm shows ineffective results when color features captured by the camera sensor are modified than that of the actual, so to diminish this, various operations within the mobile application can be provided which can enhance and improve these features of the image taken by the camera. In [31], Pathan et al. aimed to reduce the feature set, by eliminating the features which are redundant, also algorithm for thick hair removal is provided. Results can be optimized by using more efficient classifier such as SVM. Diniz and Corderio [32] modeled the concept of fuzzy numbers to perform automatic segmentation of skin lesions as skin boundaries are vague and unpredictable. The algorithm was performed using two parameters: beta and size of neighborhood mask. Experiments conducted using different values of beta and size and observed that taking beta as 0.9 and size as 11 in addition with preprocessing yields better sensitivity. Proposed algorithm has sensitivity, specificity, Jaccard index, and balanced accuracy of 0.77, 0.94, 0.65, and 0.35, respectively. Also, the model is compared to three other algorithms; apart from this, it is clear that addition of preprocessing yields higher rates. In spite of well-established results, the system is not able to classify the lesions, so need to incorporate the classifier. Etayef et al. [33] perform automatic segmentation by compounding PSO and MRF. Qualitative aspects of the proposed algorithm compared with PSO, FCM and ground truth values provided by dermatologists and quantitative analysis performed by comparing with seven other algorithms. Proposed method yields better results than others. Here also the problem is left up to segmentation stage only, so there is a place for addition of classifier.

In [34], Do et al. introduce a new concept of light and fast framework for lesion segmentation as smartphones have limited memory and rigorous computation. Apart from this, smartphone-based images are utilized which are more challenging to work with as they are sensed in normal conditions and may subject to a variety of noises. Furthermore, HCl is taken into consideration. System has a sensitivity of 89.09%, and specificity is more than 90%. Conoci et al. [35] proposed high-speed PoC device for melanoma detection. This method exploits handcrafted features for classification. Results are compared with 10 other methods, from this observed that proposed method outperformed all. This efficient approach can further be enhanced by including other skin diseases as well. Arasi et al. [36] observed the following results: ANN more efficient than ANFIS, HLF fused with SVM yields better results than LBP

fused with SVM, DWT and SVM give effective melanoma detection. ABCD parameters provide better results with ANN and ANFIS than SVM. There is a need to segment lesion so that unwanted skin lesions are removed, and this can incline the performance. Adqed et al. [37] fuse textural and structural features which are then fed to SVM for the classification. Proposed algorithm is compared to two other algorithms and observed that results are in favor of proposed algorithm. The algorithm can be extended as mobile-based application. In [38], Rajesh proposed GUI to detect melanoma, based on ABCD features and BPNN to classify images. This method can be improved by using more optimized classifier such as SVM. Lopez et al. [39] consider three methods: firstly, training CNN from scratch, secondly, VGGnet already trained net on large dataset, and thirdly fine-tuning of CNN by keeping transfer learning paradigm as such. In the first method, difference between training and testing reveals that model neither overfits nor underfits. In the second method, better results are achieved for performance parameters. In the third method, overfitting is observed. Thus, this system involves two classifiers and has sensitivity and precision of 78.66 and 79.44%. Hussain et al. [40] detect and locate tumors in rats in subcutaneous tumor transplantation models. As tumors sit beneath the skin and have same color patterns as that of the skin making the task more difficult. Pretrained VGG-16 used to detect tumor and locate tumor using SVM. This method can be employed in humans for detecting the skin lesions. Mishra and Daescu [41] use deep learning concept to segment the skin lesions from the images. Results of the proposed segmentation are compared with Otsu thresholding in which the former outperforms the latter. Results are among the top submissions of ISBI. Results of this renowned research can be extended for disease classification. Thao and Quang [42] exploit ConvNet and VGG 16 for classification and observed that results are better using latter. Average value of AUC in two tasks is 0.816. Also, these results are among top 15 submissions. Concept can be extended to other diseases as well. Comprehensive review of various state-of-art techniques for cancer detection is provided in Table 1.

3.2 Other Skin Diseases

Psoriasis is a chronic disease, characterize by abnormal patches and deposits on skin. It can affect people of any age and generally considered as of heredity background [43]. Acne is a disease which generally affects people at some stage of life. Its severity is variable ranging from small pimples to widespread bumps which generally leads to bothersome and discomfort. Erythema-squamous are the diseases which look similar in appearances but having different reasons for their origin; thus, these diseases are challenging to classify. Orofacial diseases are those which occurred within the mouth, lips, jaw, face, and tongue. Legs and feet ulcers are the wounds or cuts present on the legs and feet. These inflammations are sometimes not healed within the time and may turn into long-lasting diseases. This disease is mostly observed in elderly people. Warts are non-cancerous masses, having color same as that of the skin. Tinea Corporis leads to formation of rings on the body. It is essential to prevent

Table 1 Review of various techniques to detect skin cancer

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[14] IEEE 2017	ISIC	–	–	SIFT used to extract features such as scaling, rotation	Using Bag of features (K-means clustering)	NN-NGSA-II	Classification of basal cell carcinoma and angioma	90.56
[15] IEEE 2017	Skin Vision Picture, International skin cancer collaboration 2016	RGB to gray, contrast enhancement, histogram equalization, and median filtering	Otsu thresholding	Color, shape, TDS score, texture features and also mean skewness, kurtosis	Principal component analysis	Kernel SVM incorporated with radial basis function	Classification of malignant and Benign cancer	92.1
[16] IEEE 2017	ISIC	–	–	By employing DWT texture features extracted	PCA followed by GLCM matrix formation	NN-NGSA-II	Classification of basal cell carcinoma, skin angioma, and lentigo simplex	87.92

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[17] IEEE 2017	PH2	RGB to gray, hair removal using 2D-derivative of Gaussian and Fast Marching in-painting algorithm	Otsu and Active- Contour algorithm	Features used—2D- FFT, 2D-DCT, mean, mode, standard deviation, color, orientation, intensity pattern, variation pattern	—	SVM	Classification of benign, atypical and melanoma, Hair removal, mobile based	C1—91.5, C2—92
[18] IEEE 2017	PH2	RGB to gray	—	LBP and domain- specific features	Genetic programming	Genetic Programming	Classification of melanoma and non- melanoma	—
[19] IEEE 2017	PH2	RGB to HSV	—	Color and texture features	—	SVM	Classification of melanoma and non- melanoma	96
[20] IEEE 2017	—	RGB to grayscale, median filtering	Sobel edge detection	ABCD features	—	ANN based	Detection of melanoma stage	—

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[21] Science Direct 2015	–	Image resize, brightness and contrast adjustment with gamma correction	Otsu thresholding in each of the RGB plane then 3-plane masking and edge detection	Geometry- based features	–	ABCD rule (STOLZ's algorithm)	Classification of melanoma lesion and normal mole	–
[22] IEEE 2017	Self-collected	–	GAN's (Generative adversarial networks)	–	–	–	Pigmented and non- pigmented lesion segmentation, mobile based	92
[23] IEEE 2017	PH2	RGB to XYZ plane than extracted z channel, median filtering, followed by histogram equalization	Region growing method (using two threshold) then calculation of solidity and extent to select final image	–	–	–	Melanoma and non- melanoma lesion segmentation	–

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[24] IEEE 2017	Sydney Melanoma Diagnostic center in Royal Prince Alfred Hospital and other internet sources	Image resize, denoizing to remove hairs, air bubbles and, background noise	K-means segmentation	Wavelet transform, curvelet transform, and GLCM	—	Bayesian decision fusion (results of three SVM networks are pooled)	Classification of malignant and benign	84.1
[25] IEEE 2017	PH2 and ISIC 2016	Hair removal using Dullrazor, background detection	Saliency-based map via Bayesian network	—	—	—	Melanoma lesion segmentation	—
[26] IEEE 2017	Dermis and DermQuest	RGB to gray	K-mean clustering followed by entropy and threshold method	—	—	—	Melanoma lesion segmentation	—
[27] IEEE 2017	Archive of the department of dermatology of the university Medical Center Groningen	—	Combined K-mean clustering and Otsu segmentation	Color, texture, non-linear features, border, and asymmetry features	SVM, KNN, random forest, decision tree	Classification of malignant melanoma and benign	89.7	(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[28] IEEE 2017	Derm101, PCDS, Hosei dataset	RGB to gray, image resize	–	BEMD followed by LBPV features	AdaBoost	AdaBoost	Classification of malignant melanoma and benign	96
[29] IEEE 2017	PH2	Hair removal using in-painting method and Frangi filter	Proposed automatic threshold calculation method	Asymmetry and color features	–	Multilayer perceptron ANN (with scaled conjugate gradient algorithm)	Classification of common nevus, atypical nevus and, melanoma	–
[30] IEEE 2017	For testing-PH2 and ISDIS	Rescaling, smoothing, grayscale conversion	Edge detection, biggest contour detected and cropped	ABCD features	–	STOLZ's algorithm (ABCD rule)	Classification of malignant melanoma and benign, Mobile-based cancer	–
[31] IEEE 2017	PH2	Hair detection and removal by using bottom hat transform in combination with exemplar- based In-painting	Fuzzy c-means clustering	Shape, color variation, texture features	–	Multilayer feed forward BPNN	Classification of malignant melanoma and benign	82

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[32] IEEE 2017	ISDI	Using closing operation hairs are removed	Based on fuzzy numbers	–	–	–	Malignant and benign lesion segmentation	–
[33] IEEE 2017	PH2	Blue channel extraction, Thresholding to remove reflection artifacts and bubbles of air, then hair detection and exclusion using directional Gabor filter and in-painting	By compounding of PSO and Markov random field	–	–	–	Malignant and benign lesion segmentation	95
[34] IEEE 2018	National skin center Singapore	Downsampling	Integration of skin detection and hierarchical segmentation (using Otsu and MST method)	Color, border, asymmetry and texture features (using GLCM and LBP)	NMIFS	Four SVM classifiers	Classification of malignant melanoma and benign, mobile phone based, human-computer interface	–

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[35] IEEE 2018	PH2	RGB to YCbCr need Y channel, then resize	—	Ad-hoc customized Handcrafted features	—	Feedforward ANN (trained with BPNN)	Classification of malignant melanoma and benign	—
[36] IEEE 2017	Dermis and DermQuest	RGB to gray, bilinear interpolation closing operation for hair removal, adaptive median filtering and, contrast adjustment	—	DWT	PCA	ANN (BPNN) and ANFIS	Classification of malignant melanoma and benign	ANN—98.8 ANFIS—95.18
[37] IEEE 2017	PH2	Dullrazor for hair removal	—	Structural features (curvelet and wavelet transform), textural (using LBP)	—	SVM with linear kernel	Classification of common nevus, atypical nevus and, melanoma	86.07
[38] IEEE 2017	—	RGB to HSV	Region growing, edge detection	—	BPNN	Classification of malignant melanoma and benign	—	

(continued)

Table 1 (continued)

Work cited, published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Tasks performed	Accuracy (%)
[39] IEEE 2017	ISIC archive	Normalize to [0,1], cropping, resizing	–	–	–	Pretrained VGGNet used (transfer learning)	Classification of malignant melanoma and benign	95.95
[40] IEEE 2017	Subcutaneous xenograft taken using digital camera	Data augmentation, addition of false negatives, image resizing	–	Pretrained CNN VGG-16 and extracted multiple layers	–	SVM	Skin tumor detection	89.85
[41] IEEE 2017	ISIC archive provided by ISBI 2017	Cropping and resizing by bilinear interpolation	Unet architecture	–	–	–	Melanoma lesion segmentation	92.8
[42] IEEE 2017	ISIC 2017	Resizing, mean subtraction and normalize to [-1,1], data augmentation	Convolutional deconvolutional architecture	–	–	1. Simple ConvNet 2. Pretrained VGG16 architecture using transfer learning	Classification of melanoma skin lesion	Method 1-51.5 Method 2-83.7

the disease from spreading to other parts of body. Vitiligo is a disease in which skin loses its color and become white colored. It is also long-term disease. White patches occur because of dying out of melanocytes. Eczema skin disease can lead skin to redness, roughness, and itchiness. This disease is mostly prevalent among children.

Archana et al. [44] detected six diseases which are mostly prevalent in India, for this DCT, SWT, and SVD are used in parallel. Acne is detected only with SVD. System can be made more powerful by using more optimized classifiers. Maroni et al. [45] proposed framework, in which firstly body part is detected which can be front, left, right part of face or back. Front face detected using Haar classifier. If front face is detected, then portion of mouth blotted out as it can be misclassified; further, if left or right profile is detected, then ear portion is blotted out. Then, skin segmentation is performed followed by feature extraction and classification. Performance of six classifiers compared out of which random forest outperformed. Furthermore, heat maps are generated to focus on red areas present in the skin. Then, acne is extracted using adaptive thresholding on heat maps. In last blobs are detected using Laplacian of Gaussian filter. This method can be extended to other skin lesions apart from acne. In [46], Pasero and Castagneri perform ulcer segmentation and compared their results with the manual segmentation. This method can be used to make a record of patient's history to check whether the prescribed drugs are proving effective for patient or not. In [47], Fidan et al. classified six erythematous-squamous diseases. SOM maps are formed based on clinical and pathological data and then classification also performed using SOM. It is evident from this study that clinical data proves more fruitful in classifying diseases as compared to the pathological. George et al. [48] perform psoriasis segmentation by exploiting group of classifiers. It has been observed that this combination yields better results as compared to individual classifiers. Also, morphological operations are applied to refine the image. Furthermore, tattoo and red skin can easily be classified by this method. Problem arises as hairs and white scales which mislead the classification in most of the classifiers, and this is improved by combining the classifiers and morphological operations. This method can further be enhanced by incorporating hair removal algorithm. In [49], Anantharaman et al. identified multiple oral diseases. Mainly the classification is performed between the canker sore and cold sore. Deep learning technique is being used to train as well as classify. Effective classification technique can be utilized with OScan to perform diagnosis. Proposed system includes various layers which are user layer, web server layer, app server, and Machine learning interface. All these layers work in conjunction for providing desired results. Method can be extended to other diseases except orofacial diseases. Comprehensive review of various state-of-art techniques for other skin diseases detection is mentioned in Table 2.

4 Conclusion and Future Scope

It is evident that detection of skin diseases plays a crucial role in numerous ways. In this research, comprehensive review of various techniques used for detecting skin

Table 2 Review of various techniques to detect other diseases

Work cited published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Task performed	Accuracy (%)
[44] IEEE 2017	Dermnet	Image resizing	–	–	–	DCT, DWT, and SVD (all work in parallel)	Classification of Warts, acne, Tinea corporis, vitiligo, nail psoriasis, eczema, and mobile based	80
[45] IEEE 2017	FSD, Dermnet, and DermQuest	–	Skin segmentation performed	Color, texture, shape, spatial, unsupervised features	Mean decrease impurity (tree-based method) provided by random forest model (15 features selected)	Random forest model	Acne detection, lesion counting, and mobile based	–
[46] IEEE 2017	Self-collected from patients	Median filtering	Automatic segmentation	–	–	–	Legs and feet ulcer segmentation and mobile based	–
[47] IEEE 2016	UCI machine learning database	–	Using SOM (self-organizing map)	–	–	Using SOM	Classification of erythematous-squamous diseases	–

(continued)

Table 2 (continued)

Work cited published in and year	Database used	Preprocessing	Segmentation	Features extraction	Features selection	Classifier used	Task performed	Accuracy (%)
[48] IEEE 2017	Self-collected	–	Using 7 classifiers (SVM, KNN, NB, DT, RF, MLP and AdaBoost) based on majority voting	Pixel color values (RGB, YCbCr, HSV, XYZ, and CIE-Lab) for segmentation	–	–	Psoriasis Segmentation	97.4
[49] IEEE 2017	http://images.google.com	–	–	–	–	Using 1. Clarifai algorithm 2. Random forest 3. Pretrained Inception v3 model	Classification of Orofacial diseases (cold sore and canker sore) and mobile app based	1. 100% 3. 66%

diseases is carried out. Diseases included are skin cancers, acne, psoriasis, orofacial diseases, erythematous-squamous diseases, warts, tinea corporis, vitiligo, eczema, legs, and feet ulcers. Also their comparative analysis is made, which presents the data in concise form. It is clear that in case of skin cancer detection, deep learning techniques yield reliable results. Best results are achieved when VGGnet CNN is exploited using transfer learning. Other than this, it is also observed that recent work is inclined more toward skin cancer detection and classification while less toward other diseases detection. Further, it has been seen that the number of diseases that can be detected using CAD systems is less, so there can be addition of more diseases to be detected thus making the system more diverse and fruitful in the field of dermatology. Skin disease detection is a challenging task, so it can be clearly seen that there is still a scope for improvement for efficient skin disease detection for making an effective CAD tool.

References

1. D'Orazio, J., Jarrett, S., Amaro-Ortiz, A., Scott, T.: UV radiation and the skin. *Int. J. Mol. Sci.* **14**(6), 12222–12248 (2013). <https://doi.org/10.3390/ijms140612222>
2. Proksch, E., Brandner, J.M., Jensen, J.: The skin: an indispensable barrier. *Exp. Dermatol.* **17**(12), 1063–1072 (2008). <https://doi.org/10.1111/j.1600-0625.2008.00786.x>
3. NUSKIN.https://www.nuskin.com/en_ZA/corporate/company/science/skin_care_science/skin_anatomy_andphysiology.html
4. Health. <http://www.health.com/skin-conditions>
5. Hollestein, L.M., Nijsten, T.: An insight into the global burden of skin diseases. *J. Invest. Dermatol.* **134**(6), 1499–1501 (2014). <https://doi.org/10.1038/jid.2013.513>
6. Hay, R.J., Johns, N.E., Williams, H.C., Bolliger, I.W., Dellavalle, R.P., Margolis, D.J., Marks, R., Naldi, L., Weinstock, M.A., Wulf, S.K., Michaud, C., Murray, C.J.L., Naghavi, M.: The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. *J. Invest. Dermatol.* **134**(6), 1527–1534 (2013). <https://doi.org/10.1038/jid.2013.446>
7. Lim, H.W., Collins, S.A.B., Resneck, J.S., Bologna, J.L., Hodge, J.A., Rohrer, T.A., Van Beek, M.J., Margolis, D.J., Sober, A.J., Weinstock, M.A., Nerenz, D.R., Begolka, W.S., Moyano, J.V.: The burden of skin disease in the United States. *J. Am. Acad. Dermatol.* **76**(5), 958–972.e2 (2017). <https://doi.org/10.1016/j.jaad.2016.12.043>
8. Jain, S., Barambhe, M.S., Jain, J., Jajoo, U.N., Pandey, N.: Prevalence of skin diseases in rural Central India: a community-based, cross-sectional, observational study. *J. Mahatma Gandhi Inst. Med. Sci.* **21**, 111–115 (2016). <https://doi.org/10.4103/0971-9903.189537>
9. Manoorkar, P.B., Kamat, D.K., Patil, P.M.: Analysis and classification of human skin diseases. In: International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), pp. 1067–1071. IEEE Press, Pune (2016). <https://doi.org/10.1109/icacdot.2016.7877750>
10. Zogkas, A., Gilou, S., Kirsanidou, I., Korfitis, C., Kemanetzi, C., Lazaridou, E., Panagiotis, B., Astaras, A.: A non-invasive medical decision support prototype system for dermatology based on electrical impedance spectroscopy (dermasense). In: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pp. 195–199. IEEE Press, Thessaloniki (2017). <https://doi.org/10.1109/cbms.2017.86>
11. Mirbeik-Sabzevari, A., Ashinoff, R., Tavassolian, N.: Ultra-wideband millimeter-wave dielectric characteristics of freshly-excised normal and malignant human skin tissues. In: IEEE Transactions on Biomedical Engineering, pp. 1320–1329. IEEE Press (2017). <https://doi.org/10.1109/tbme.2017.2749371>

12. Lacy, K., Wisam, A.: Skin cancer. Medicine **41**(7), 402–405 (2013). <https://doi.org/10.1016/j.mpmed.2013.04.008>
13. Farberg, A.S., Rigel, D.S.: The importance of early recognition of skin cancer. Dermatol. Clin. **35**(4), 15–16 (2017). <https://doi.org/10.1016/j.det.2017.06.019>
14. Chakraborty, S., Mali, K., Chatterjee, S., Anand, S., Basu, A., Banerjee, S., Das, M., Bhattacharya, A.: Image based skin disease detection using hybrid neural network coupled bag-of-features. In: IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 242–246. IEEE Press, New York (2017). <https://doi.org/10.1109/uemcon.2017.8249038>
15. Alquran, H., Qasmieh, I.A., Alqudah, A.M., Alhammouri, S., Alawneh, E., Abughazaleh, A., Hasayen, F.: The melanoma skin cancer detection and classification using support vector machine. In: IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), pp. 1–5. IEEE Press, Aqaba (2017). <https://doi.org/10.1109/aeect.2017.8257738>
16. Chakraborty, S., Mali, K., Chatterjee, S., Banerjee, S., Mazumdar, K.G., Debnath, M., Basu, P., Bose, S., Roy, K.: Detection of skin disease using metaheuristic supported artificial neural networks. In: 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), pp. 224–229. IEEE Press, Bangkok (2017). <https://doi.org/10.1109/iemecon.2017.8079594>
17. Joseph, S., Panicker, J.R.: Skin lesion analysis system for melanoma detection with an effective hair segmentation method. In: International Conference on Information Science (ICIS), pp. 91–96. IEEE Press, Kochi (2016). <https://doi.org/10.1109/infosci.2016.7845307>
18. UIAin, Q., Xue, B., Al-Sahaf, H., Zhang, M.: Genetic programming for skin cancer detection in dermoscopic images. In: IEEE Congress on Evolutionary Computation (CEC), pp. 2420–2427. IEEE Press, San Sebastian (2017). <https://doi.org/10.1109/cec.2017.7969598>
19. Waheed, Z., Waheed, A., Zafar, M., Riaz, F.: An efficient machine learning approach for the detection of melanoma using dermoscopic images. In: 2017 International Conference on Communication, Computing and Digital Systems (C-CODE), pp. 316–319. IEEE Press, Islamabad (2017). <https://doi.org/10.1109/c-code.2017.7918949>
20. Reshma, M., Shan, B.P.: Two methodologies for identification of stages and different types of melanoma detection. In: Conference on Emerging Devices and Smart Systems (ICEDSS), pp. 257–259. IEEE Press, Tiruchengode (2017). <https://doi.org/10.1109/icedss.2017.8073689>
21. Jain, S., Jagtap, V., Pise, N.: Computer Aided Melanoma Skin Cancer Detection Using Image Processing. In: International Conference on Computer, Communication and Convergence (ICCC 2015), pp. 735–740. Procedia Computer Science (2015). <https://doi.org/10.1016/j.procs.2015.04.209>
22. Udrea, A., Mitra, G.D.: Generative adversarial neural networks for pigmented and non-pigmented skin lesions detection in clinical images. In: 21st International Conference on Control Systems and Computer Science (CSCS), pp. 364–368. IEEE, Bucharest (2017). <https://doi.org/10.1109/cscs.2017.56>
23. Agarwal, A., Issac, A., Dutta, M.K.: A region growing based imaging method for lesion segmentation from dermoscopic images. In: 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), pp. 632–637. IEEE Press, Mathura (2017). <https://doi.org/10.1109/upcon.2017.8251123>
24. Takruri, M., Abubakar, A. Bayesian decision fusion for enhancing melanoma recognition accuracy. In: International Conference on Electrical and Computing Technologies and Applications (ICECTA), pp. 1–4. IEEE Press, Ras Al Khaimah: (2017). <https://doi.org/10.1109/icecta.2017.8252063>
25. Ahn, E., Kim, J., Bi, L., Kumar, A., Li, C., Fulham, M., Feng, D.D.: Saliency-based lesion segmentation via background detection in dermoscopic images. IEEE J. Biomed. Health Inform. 1685–1693 (2017). <https://doi.org/10.1109/jbhi.2017.2653179>
26. Patel, B., Dhayal, K., Roy, S., Shah, R.: Computerized skin cancer lesion identification using the combination of clustering and entropy. In: International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), pp. 46–51. IEEE Press, Chirala (2017). <https://doi.org/10.1109/icbdaci.2017.8070807>

27. Munia, T.T.K., Alam, M.N., Neubert, J., Fazel-Rezai, R.: Automatic diagnosis of melanoma using linear and nonlinear features from digital image. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 4281–4284. IEEE Press, Seogwipo (2017). <https://doi.org/10.1109/embc.2017.8037802>
28. Fonseca-Pinto, R., Machado, M.: A textured scale-based approach to melanocytic skin lesions in dermoscopy. In: 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 279–282. IEEE Press, Opatija (2017). <https://doi.org/10.23919/mipro.2017.7973434>
29. Turkeli, S., Oguz, M.S., Abay, S.B., Kumbasar, T., Atay, H.T., Kurt, K.K.: A smart dermoscope design using artificial neural network. In: International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1–6. IEEE Press, Malatya (2017). <https://doi.org/10.1109/idap.2017.8090211>
30. Firmansyah, H.R., Kusumaningtyas, E.M., Hardiansyah, F.F.: Detection melanoma cancer using ABCD rule based on mobile device. In: International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), pp. 127–131. IEEE Press, Surabaya (2017). <https://doi.org/10.1109/kcic.2017.8228575>
31. Pathan, S., Siddalingaswamy, P.C., Lakshmi, L., Prabhu, K.G.: Classification of benign and malignant melanocytic lesions: a CAD tool. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1308–1312. IEEE Press, Udupi (2017). <https://doi.org/10.1109/icacci.2017.8126022>
32. Diniz, J.B., Cordeiro, F.R.: Automatic segmentation of melanoma in dermoscopy images using fuzzy numbers. In: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pp. 150–155. IEEE Press, Thessaloniki (2017). <https://doi.org/10.1109/cbms.2017.39>
33. Eltayef, K., Li, Y., Liu, X.: lesion segmentation in dermoscopy images using particle swarm optimization and Markov Random Field. In: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS), pp. 739–744. IEEE Press, Thessaloniki (2017). <https://doi.org/10.1109/cbms.2017.26>
34. Do, T.T., Hoang, T., Pomponiu, V., Zhou, Y., Chen, Z., Cheung, N.M., Koh, D., Tan, A., Hoon, T.: Accessible melanoma detection using smartphones and mobile image analysis. IEEE Trans. Multimedia (2018). <https://doi.org/10.1109/tmm.2018.2814346>
35. Conoci, S., Rundo, F., Petralta, S., Battiatto, S.: Advanced skin lesion discrimination pipeline for early melanoma cancer diagnosis towards PoC devices. In: European Conference on Circuit Theory and Design (ECCTD), pp. 1–4. IEEE Press, Catania (2017). <https://doi.org/10.1109/ecctd.2017.8093310>
36. Arasi, M.A., El-Horbaty, E.S.M., Salem, A.M., El-Dahshan, E.S.A.: Computational intelligence approaches for malignant melanoma detection and diagnosis. In: 8th International Conference on Information Technology (ICIT), pp. 55–61. IEEE Press, Amman (2017). <https://doi.org/10.1109/icitech.2017.8079915>
37. Adjed, F., Safdar Gardezi, S.J., Ababsa, F., Faye, I., Chandra Dass, S.: Fusion of structural and textural features for melanoma recognition, pp. 185–195. In: IET Computer Vision. IET Press (2018). <https://doi.org/10.1049/iet-cvi.2017.0193>
38. Rajesh, A.: Classification of malignant melanoma and Benign Skin Lesion by using back propagation neural network and ABCD rule. In: IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), pp. 1–8. IEEE Press, Karur (2017). <https://doi.org/10.1109/iceice.2017.8191916>
39. Lopez, A.R., Giro-i-Nieto, X., Burdick, J., Marques, O.: Skin lesion classification from dermoscopic images using deep learning techniques. In: 13th IASTED International Conference on Biomedical Engineering (BioMed), pp. 49–54. IEEE Press, Innsbruck (2017). <https://doi.org/10.2316/p.2017.852-053>
40. Hussain, N., Cooper, P., Shnyder, S., Ugail, H., Bukar, A.M., Connah, D.: A non-invasive 2D digital imaging method for detection of surface lesions using machine learning. In: International Conference on Cyberworlds (CW), pp. 166–169. IEEE Press, Chester (2017). <https://doi.org/10.1109/cw.2017.39>

41. Mishra, R., Daescu, O.: Deep learning for skin lesion segmentation. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1189–1194. IEEE Press, Kansas (2017). <https://doi.org/10.1109/bibm.2017.8217826>
42. Thao, L.T., Quang, N.H.: Automatic skin lesion analysis towards melanoma detection. In: 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES), pp. 106–111. IEEE Press, Hanoi (2017). <https://doi.org/10.1109/iesys.2017.8233570>
43. MedicineNet. https://www.medicinenet.com/psoriasis/article.htm#psoriasis_facts
44. Ajith, A., Goel, V., Vazirani, Roja, M.M.: Digital dermatology: skin disease detection model using image processing. In: International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 168–173. IEEE Press, Madurai (2017). <https://doi.org/10.1109/iccons.2017.8250703>
45. Maroni, G., Ermidoro, M., Previdi, F., Bigini, G.: Automated detection, extraction and counting of acne lesions for automatic evaluation and tracking of acne severity. In: IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–6. IEEE Press, Honolulu (2017). <https://doi.org/10.1109/ssci.2017.8280925>
46. Pasero, E., Castagneri, C.: Application of an automatic ulcer segmentation algorithm. In: IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI), pp. 1–4. IEEE Press, Modena (2017). <https://doi.org/10.1109/ssci.2017.8280925>
47. Fidan, U., Ozkan, N., Calikusu, I.: Clustering and classification of dermatologic data with Self Organization Map (SOM) method. In: Medical Technologies National Congress (TIPTEKNO), pp. 1–4. IEEE Press, Antalya (2016). <https://doi.org/10.1109/tiptekno.2016.7863075>
48. George, Y., Aldeen, M., Garnavi, R.: A pixel-based skin segmentation in psoriasis images using committee of machine learning classifiers. In: International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8. IEEE Press, Sydney (2017). <https://doi.org/10.1109/dicta.2017.8227398>
49. Anantharaman, R., Anantharaman, V., Lee, Y.: Oro Vision: deep learning for classifying orofacial diseases. In: IEEE International Conference on Healthcare Informatics (ICHI), pp. 39–45. IEEE Press, Park City (2017). <https://doi.org/10.1109/ichi.2017.69>

A Green Computing Technique for Reducing Energy Consumption for Desktops Using Proxy Downloading



Meenakshi S. Arya, Ashwini Manjare, Payal Naik and Kamrul Haq Sheikh

Abstract Computers extensively consume chemicals, fossil fuels and water for their manufacture. When put to use, they are also a prominent source of carbon dioxide emission, emitting about 0.1 tons of carbon dioxide in a year. There is a significant amount of difference in the energy consumption of a computer depending upon whether the screen is powered on or off. Due to the inherent problem of Internet connection getting hampered while switching the machines to sleep mode, the idle desktop machines are seldom switched to sleep mode in enterprise networks. The paper proposes a system for reducing energy consumption and emission of carbon dioxide by migrating file downloading from host machine to server. The same is achieved by switching host machines to sleep mode whenever they are idle. Furthermore, an effort to download an existing file which might have been downloaded already at server will be cancelled to reduce redundant network traffic. The system does not require any hardware additions to be done to the existing systems. The software needs to be directly installed onto the system.

Keywords Energy saving · CO₂ Emission · Performance enhancement · Local area network · Green desktop computing

1 Introduction

With the emphasis being laid on making earth a greener planet, significant attention during recent times has been diverted towards the energy consumed by the computing infrastructure. A study conducted in [1] concludes that PCs and more specifically the display units consume about 100 TWh/year, and nearly 65% of the same is being utilized by desktop computers in enterprises. At present, energy expenditure

M. S. Arya (✉) · P. Naik · K. H. Sheikh
G H Raisoni College of Engineering, Nagpur, India
e-mail: meenakshi.arya@raisoni.net

A. Manjare
Vidyalankar Institute of Technology, Mumbai, India

in PC mainly lies on the fact that it is universal to let machines enter a low-power sleep state; however, during that state as well, there is not significant conservation of energy. The study was a pioneering one with its main purpose being supplementing and updating of earlier data that was collected to determine the amount to which equipment at workplace is either switched off or automatically switches over to a low-power state on being inactive for a specific period of time. The data provided improved the assessment of how many units of energy are typically consumed and how much savings are being done for different kinds of equipment type, thereby guiding the ENERGY STAR Office Equipment program to lay emphasis on and paving a way for their effort on development of future products with the maximum energy savings capabilities.

The study also emphasizes that 67% of the desktop computers continue to stay on a power on mode beyond work hours, and merely 4% switch to sleep mode. The basis of the same has been discovered and validated in [2], which concludes that nearly 40% of such instances are meant for file downloading. Turning off the idle system increases the likelihood of the Internet connection to break especially during file downloading. At the same time, downloading redundant file (through repeated download of the same file over a LAN) also causes wastage of energy making the PC an energy-consuming monster.

The proposed system has two aims: First, when the host turns to idle condition or the user leaves for a certain duration, the system ought to be capable of switching the host to sleep mode, and in case there are files being downloaded, the network connectivity should not be interrupted thereby allowing the file downloading even to continue despite the host being in sleep mode. Second, downloading of an already existing file in the same LAN (which might have been downloaded earlier) should not be permitted to continue in order to decrease the redundant network traffic. By switching host machines to sleep mode, it also reduces the emission of CO₂.

The paper is structured as follows. Section 2 presents the literature survey while Sect. 3 gives a description of the proposed method; the outcomes are discussed in Sect. 4. Section 5 summarizes and concludes the proposed work.

2 Related Work

The proposed algorithm in [3] does not interrupt the file downloading despite the PC switching to sleep mode. The proposed system also causes a significant reduction in the network traffic; however, the biggest drawback is that the system is capable of downloading only http and ftp files and is platform dependent as it works in Linux OS only.

The authors in [4] proposed a virtualization-based energy saving system. The user's desktop computing environment is virtualized as a virtual machine (VM), and during the idle state of the desktop computing environment, it is migrated to the same. Thus, the environment of the user's desktop remains potentially on, even though the physical machine is in energy saving mode and has turned itself off. This

system ensures that the ongoing connections and the running applications remain undisturbed through encapsulation of the total desktop computing environment in a VM. However, the VM software must be installed by the user to implement this approach.

A hardware-based solution for green computing is proposed in [2]. They supplement the Network Interface Card with a low-powered hardware device as compared to the common PC. The supplemented device takes over once the host system switches to the sleep mode. File download handling in place of the host PC is done by this additional hardware. However, the practicality of the system is a big concern as additional hardware is required to be present on every system, and modifications of applications as well as the host operating system are mandatory. These are a considerable obstacle to the pervasive implementation of this technology. In comparison, the proposed method does not require any extra hardware or alteration of any application on the client.

In [5–9], network connectivity proxy (NCP) can handle some basic network presence and management protocols like ICMP, DHCP and ARP. on behalf of sleeping hosts and wake them up only when it is truly essential. But of self-proxying, NCP is located within the network host on its NIC. Since NIC has limited memory and computational power, it will be difficult to implement a complete stand-alone proxy covering many applications.

3 System Architecture

The proposed system comprises of two lightweight software, namely cProxy (client-side Proxy) and sProxy (server-side Proxy) which are implemented at client side and server side, respectively. cProxy is responsible for detecting the user(client) downloading behaviour and the status of the host(the machine being used by the client). At the same time, a communication is established with sProxy with the aim of transmission of information. Whenever the host changes its status, the information is immediately intimated to sProxy.

All the hosts having file downloading requirements in the same LAN will be handled by sProxy, and it will process all the downloading operations on behalf of the host implying that all the task of downloading will be transferred to the server and completed by sProxy. The system architecture has been described in Fig. 1.

4 System Design

The flow of this system is as follows:

- The trigger for a file download by the user prompts the cProxy to fetch the required web link of the file under consideration. This information is further forwarded to

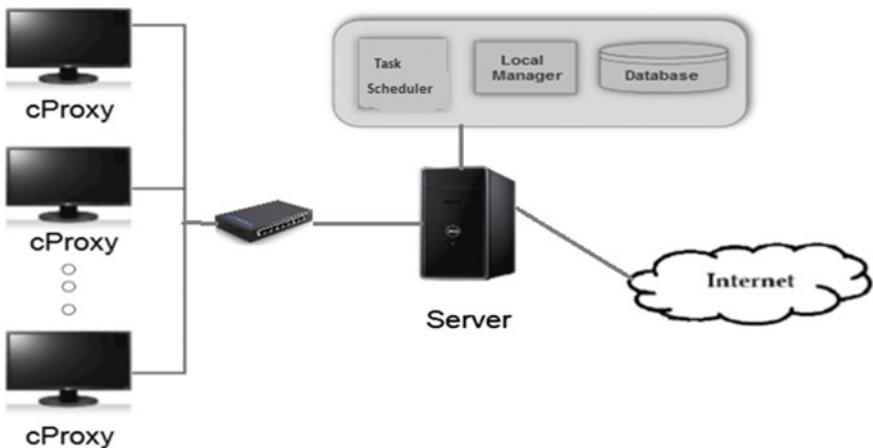


Fig. 1 System architecture

sProxy in the same subnet. Subsequently, the host can switch to sleep mode in case the user remains inactive beyond a certain period of time.

- On receipt of a requirement for file downloading from any host, sProxy initially analyses the arguments in the uniform resource locator (URL) and determine the filename. The database is queried on the basis of the file name and link, to decide whether the required file was cached earlier. In case it is not, sProxy invokes the equivalent download software to download the file in place of the host.
- On completion of the file downloading, sProxy along with sending the file back to the host once it is active (the one who had requested for the downloading of it) stores a copy of it in the cache as well for ease of access later on.
- In case the host is not active, the sProxy waits for the host to change its state and sends the file only when it receives a notice regarding a change in the status of the host.
- On receipt of a request from another host regarding downloading of a file which has been already downloaded, the requirement instructions are dispatched to sProxy.
- sProxy detects the redundant download of the file initiated by other host and cancels the fresh attempt to download the file from Internet, fetches the file from the cache where a copy is stored and sends the file back to the requesting host immediately.

5 Results and Analysis

cProxy Login: When client logs into the system using username and password (Figs. 2 and 3).

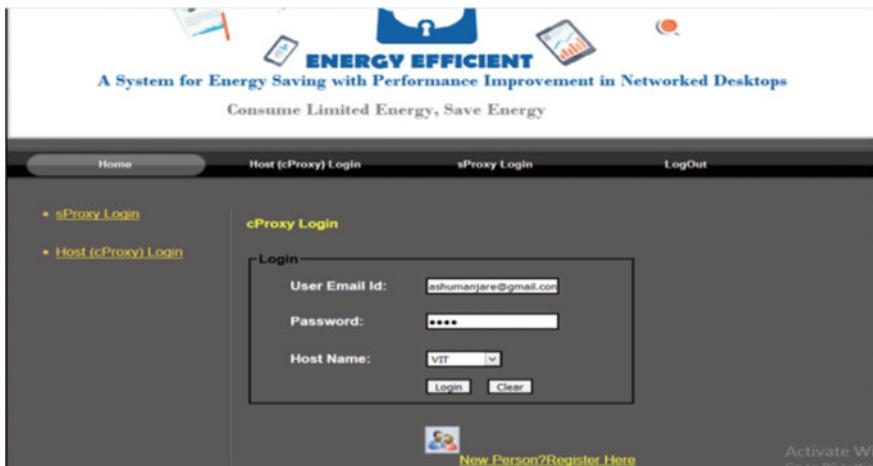


Fig. 2 Client login console

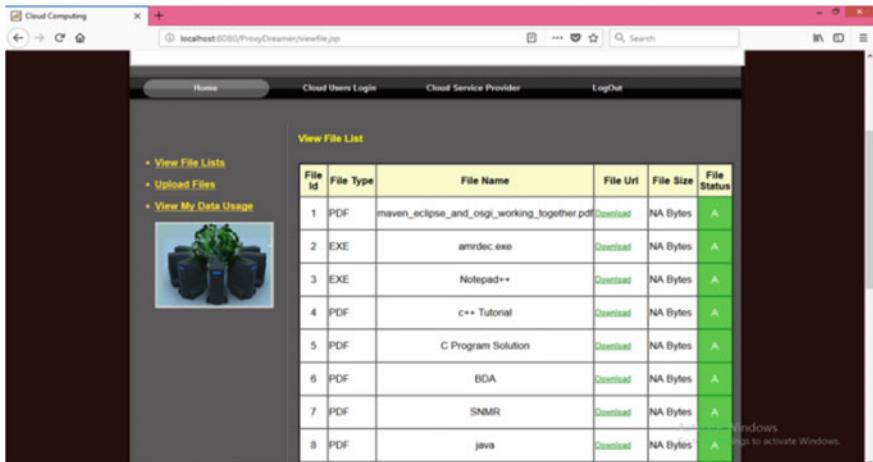


Fig. 3 Available files for downloading

When users login to system as cProxy, they can see all available files to download (Figs. 4, 5, 6, 7 and 8).

$$\begin{aligned}
 \text{Total Powersaved} &= \text{power consumed in active} - \text{power consumed in sleep mode} \\
 &= 21.64957265 - 16.17511737 \\
 &= 5.474455279 \text{ W.}
 \end{aligned}$$

Fig. 4 File downloading in progress

6 Conclusion

The technique mainly focuses on reducing energy consumption and reducing emission of CO₂. Whenever the host switches to idle status or the user remains inactive beyond a specified period of time, the system switches the host to sleep mode. The proposed technique ensures that file downloading is neither stalled nor interrupted even when the host is in sleep mode. An attempt to download an existing file will be avoided, thereby improving the download speed along with achieving the benefit of energy saving for desktop PCs. This system can be realized using two systems; however, it can be easily scaled to larger organizations.

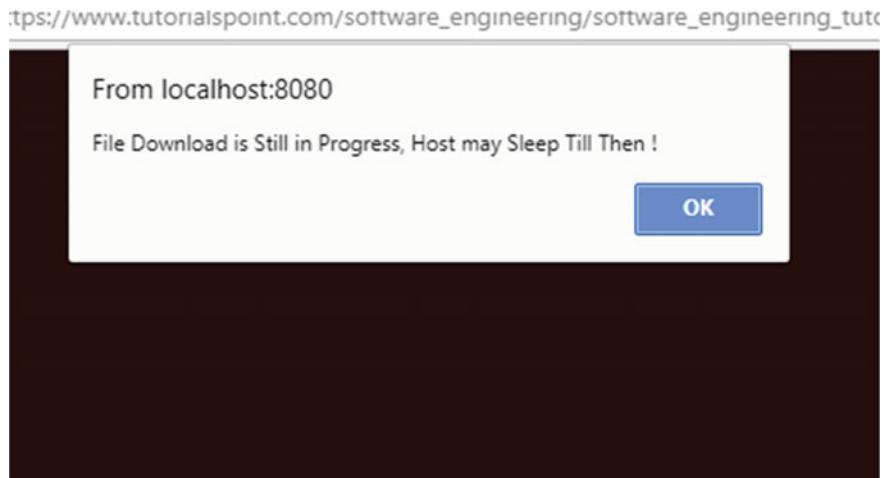


Fig. 5 File downloading is still in progress message

```
lenovo
System is not Idle
System is not Idle
System is not Idle
System is not Idle
System is idle0
System is idle1
System is idle2
System is idle3
System is idle4
http://localhost:8080/ProxyDreamer/statusinsert.jsp?username=lenovo&status=idle&count=5
Serverz Updation Status:

SUCCESS

System is not Idle
System is idle0
System is idle1
System is idle2
System is idle3
System is idle4
http://localhost:8080/ProxyDreamer/statusinsert.jsp?username=lenovo&status=idle&count=5
Serverz Updation Status:
```

Fig. 6 Idle status of user

System Id	Status
LAPT0P-LNQ59E1	idle
ashwini	idle
Rahate-PC	idle
Rahate-PC	idle
lenovo	idle

Fig. 7 User Idle status entry at sProxy

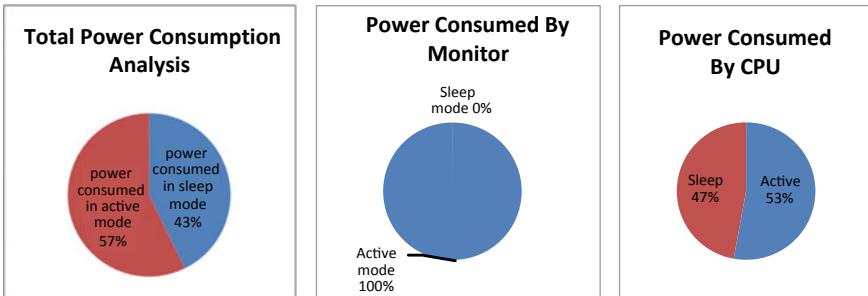


Fig. 8 Total power consumption analysis

Acknowledgements We would like to take this opportunity to express our profound gratitude and deep regard to our guide Dr. Meenakshi Arya for her perfect guidance, valuable feedback and steady support throughout the duration of the project. Her valuable suggestions were a huge help all the way through this project work. Working under her was an exceptionally educating experience for us.

References

1. Roberson, J.A., Webber, C.A., McWhinney, M.C., Brown, R.E., Pinckard, M., Busch, J.F.: After-hours power status of office equipment and energy use of miscellaneous plug-load equipment (2004)
2. Agarwal, Yuvraj, Hodges, Steve, Chandra, Ranveer, Scott, James, Bahl, Paramvir, Gupta, Rajesh: Somniloquy: augmenting network interfaces to reduce PC energy usage. NSDI **9**, 365–380

- (2009)
- 3. Zhong, W., Shi, G., Zhao, Z., Xia, F.: Parasite: a system for energy saving with performance improvement in networked desktops. In: Proceedings of the 2011 IEEE/ACM International Conference on Green Computing and Communications, pp. 79–84. IEEE Computer Society (2011)
 - 4. Das, T., Padala, P., Padmanabhan, V.N., Ramjee, R., Shin, K.G.: LiteGreen: saving energy in networked desktops using virtualization. In: USENIX annual technical conference, pp. 26–34 (2010)
 - 5. Kazandjieva, M., Heller, B., Levis, P., Kozyrakis, C. : Energy dumpster diving. In: Proceedings of 2nd Workshop on Power Aware Computing (Hot-Power'09), pp. 1–5 (2009). Khan, R., Bolla, R., Repetto, M., Bruschi, R., Giribaldi, M.: Smart proxying for reducing network energy consumption (2012)
 - 6. Khan, R., Khan, S.U.: Achieving energy saving through proxying applications on behalf of idle devices. Procedia Comput Sci 83, 187–194(2016)
 - 7. Bolla, R., Giribaldi, M., Khan, R., Repetto, M.: Design and implementation of cooperative network connectivity proxy using universal plug and play. In: The future internet assembly, pp. 323–335. Springer, Berlin (2013)
 - 8. Jimeno, M., Christensen, K., Nordman, B.: A network connection proxy to enable hosts to sleep and save energy. In: Performance, computing and communications conference, 2008. IPCCC 2008. IEEE International, pp. 101–110. IEEE (2008)
 - 9. Bolla, R., Giribaldi, M., Khan, R., Repetto, M.: Network connectivity proxy: an optimal strategy for reducing energy waste in network edge devices. In: Digital Communications-Green ICT (TIWDC), pp. 1–6. IEEE (2013)

Enhancement of Brain MR-T1/T2 Images Using Mathematical Morphology



Anu Arya, Vikrant Bhateja, Mansi Nigam and Anuj Singh Bhaduria

Abstract Brain tumor is a life-threatening disease with a fast growth rate, which makes its detection a critical task. However, low contrast and high noise content in brain MR images hamper the screening of tumor. Enhancement is therefore done to improve the perceivable features of these images. This paper presents an improved enhancement technique of brain MR-T1/T2 images by employing morphological filters. In this method, a disk-shaped flat structuring element along with top-hat and bottom-hat morphological operators is used. The performance of the filter is validated by incrementing values of contrast improvement index (CII) and peak signal-to-noise ratio (PSNR) parameters indicating a successful enhancement without noise amplification.

Keywords Bottom-hat transform · CII · Enhancement · Morphological filtering

1 Introduction

Brain is the most complicated, yet very important part of human body. It suffers from a number of abnormalities among which brain tumor has been the most ominous and intractable disease which can even cause death. It is the growth of aberrant cells in the brain that can be classified into Benign (Grade I and Grade II) or Malignant (Grade III and Grade IV) [1–3] tumors. Due to complex brain structure and varied

A. Arya (✉) · V. Bhateja · M. Nigam · A. S. Bhaduria

Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow 226028, UP, India

e-mail: 21anuarya@gmail.com

V. Bhateja

e-mail: bhateja.vikrant@gmail.com

M. Nigam

e-mail: mansi2996@gmail.com

A. S. Bhaduria

e-mail: singh.anuj842@gmail.com

impact on different individuals, screening of tumor becomes a challenging task with simple imaging techniques. Hence, magnetic resonance image (MRI) which is the most compliant and resourceful technique preferred currently is used and contains many imaging modalities such as MR-T1/T2-weighted sequences, each having its own contrast characteristics [4]. MR-T1-weighted sequences have low signal intensity for cerebrospinal fluid (CSF) with contrasting high signal intensity for MR-T2-weighted sequences [5]. The complete process of identification and detection of tumor is time-taking and laborious task [6]. Various challenges such as low-intensity contrast, unclear boundaries and background noise are always associated with MR images which might significantly affect the detection process. Therefore, it is very important to enhance the images before being processed further. A number of enhancement techniques are used amongst which adaptive histogram equalization (AHE) is very common [7]. However, it has a limitation of over-amplifying noise. So, a combination of contrast-limited adaptive histogram equalization (CLAHE) and discrete wavelet transform (DWT) was deployed by Lidong et al. [8]. In their method, only low-frequency components were enhanced to avoid noise intensification. The method preserved image details with noise suppression but did not provide a solution for alteration of image brightness. Histogram techniques tend to change the brightness of image and saturate it. Overcoming this problem, fuzzy adaptive histogram equalization technique was proposed by Magudeeswaran et al. [9]. The method was performed in three stages resulting in an enhanced image retaining original brightness values to maximum extent. Kharrat et al. [10] utilized mathematical morphology to increase MR image contrast with a disk-shaped structuring element. Benson et al. [11] proposed a method for MR image contrast enhancement and skull stripping using morphological filters. The algorithm was simple and could be applied in various applications such as tumor detection, volume analysis, and classification as reported. Hassanpour et al. [12] used morphological operations to devise a filter using top-hat transform with a disk-structuring element. The method was highly potential to enhance poor quality MR images and was simple with less processing time. To combat upon the existing challenges observed for the histogram and morphological filters toward contrast enhancement, an improved method based on the combination of top-hat and bottom-hat operations using disk-structuring element is presented in this paper. A disk-shaped structuring element is used for brain MR-T1/T2 images because it is independent of rotational changes [12]. The rest of the paper is organized in the following sections: Sect. 2 describes the general overview of morphological filtering along with the proposed methodology; Sect. 3 discusses image quality assessment, summarizes the results, performance metrics, and discusses the outcomes of this work. Section 4 summarizes the conclusion of the work.

2 Proposed Enhancement Methodology

2.1 Overview of Morphological Filtering

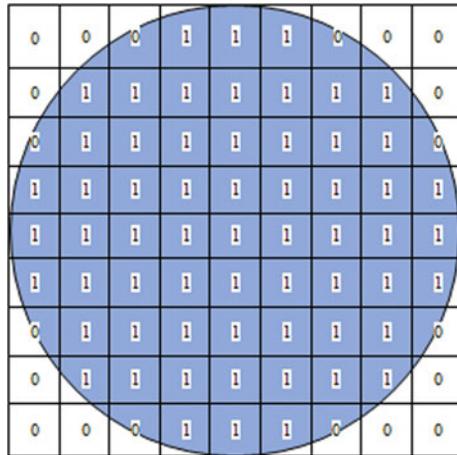
Morphological filtering [13, 14] is a set of nonlinear filtering techniques that are based on the structural properties of the objects. They are implemented through a combination of morphological operators which are related to the shape of entities in the image. These operators are applied on two inputs which are the input image and the structuring element. Structuring element is a template that designates the pixel neighborhood, and the choice of its shape and size depends upon the application and information being fetched [15]. The common morphological operators are erosion, dilation, opening, and closing [13, 16]. Erosion shrinks the image scaling down its area, whereas dilation expands the image leaving it with an increased area [17]. Opening is erosion followed by dilation, and closing is dilation followed by erosion. In the same way, these operators can further be used in a number of combinations for a variety of filtering operations. Morphological filtering finds a wide range of application in medical imaging including thermographic, magnetic resonance, and ultrasound images. In MRI, noise reduction and contrast enhancement are the primary requisites in any of its application for medical diagnosis. Hence, morphological filtering techniques which are simple to implement are very useful for MR images. However, morphological filtering poses certain limitations which make their usage a challenging task. Firstly, the results of filtering may not be very good in case of highly intricate images, and secondly, the type and size of structuring element [18] need to be chosen with utmost care, lack of which may lead to distortion of image features.

2.2 Proposed Morphological Filter for MR-T1/T2 Enhancement

This paper proposes an improved and computationally simple enhancement technique for brain MRI using morphological filtering. The implemented filter is a mathematical combination of top-hat and bottom-hat operators [12] which in terms are obtained by a combination of image subtraction with opening and closing [19, 20]. In the first stage, the input image (f) is taken in grayscale format and structuring element (x) is defined. The size of element chosen must be enough to trap tumor region without blurring the images. Disk-shaped flat structuring element [21] of size 35 as shown in Fig. 1 is suitable for this filter as size less than this trapped other small irrelevant features. Then, the image is subjected to top-hat ($T_{\text{hat}}(f)$) and bottom-hat ($B_{\text{hat}}(f)$) transformations defined by Eqs. (1) and (2), respectively.

$$T_{\text{hat}}(f) = f - (f \circ x) \quad (1)$$

Fig. 1 Disk-shaped structuring element of size 35 (the dimension of this figure scaled down to one by tenth)



$$B_{\text{hat}}(f) = (f \bullet x) - f \quad (2)$$

Top-hat transform acts as a high-pass filter and highlights the bright objects on a dark background. This transform leads to a distinct visibility of tumor region which originally was merged with the background. The most important utility of top-hat is to correct the non-uniformity in intensity which is a pervasive problem in MR images.

Bottom-hat transform is just the converse of top-hat transform and is used to highlight the darker regions of the image. Since it is desirable to minimize the background features that are not relevant for diagnostic purposes, bottom-hat transform is subtracted from top-hat in the final step denoted by Eq. (3).

$$I(f) = T_{\text{hat}}(f) - B_{\text{hat}}(f) \quad (3)$$

where $I(f)$ is the output image.

Algorithm 1: Procedural Steps for Proposed Enhancement Filter using Morphological Operators

BEGIN

Step 1: Input MR-T1/T2 image in $[I]$.

Step 2: Process Convert $[I]$ from RGB to Grayscale $[f]$.

Step 3: Define Structuring element (x) of size 35.

Step 4: Compute Top-hat transform of grayscale image $[f]$ using Eq. (1).

Step 5: Compute Bottom-hat transform of grayscale image $[f]$ using Eq. (2).

Step 6: Output Subtract Step 4 and Step 5 using Eq. (3).

END

3 Results and Discussions

3.1 *Image Quality Assessment*

Enhancement refers to the adjustment of image by modifying the values of pixels. However, it may result in amplification of noise level also. Therefore, contrast improvement cannot be considered solely for quality assessment of the enhanced image. Thus, the evaluation is done on the basis of two parameters which are contrast improvement index (CII) and peak signal-to-noise ratio (PSNR) [22, 23]. CII is used for calculating the contrast improvement, whereas noise level can be measured using PSNR. CII is the ratio of contrast of enhanced image to the contrast of original image which means that higher the value of CII more is the contrast improvement of the enhanced image. PSNR is an absolute quantity used for measuring the noise level in an image. Higher value of PSNR of the enhanced image compared to the original image indicates that the enhancement technique does not amplify noise.

3.2 *Experimental Results*

The MR images used for simulation carried out in the present work are taken from two publicly available databases: The Whole Brain Atlas [24] and Internet Brain Segmentation Repository [25]. As a preprocessing step, the image is converted from RGB to grayscale profile. The structuring element chosen was a disk of size 35 as shown in Fig. 1. Then, Eqs. (1) and (2) were applied separately on the grayscale image for top-hat and bottom-hat transformation, respectively, followed by Eq. (3) which represent the proposed filtering technique. The output image is enhanced image with distinct regions. The result of the simulations with original and enhanced images is shown in Fig. 2.

The quality assessment of images has been conducted using CII and PSNR described in Sect. 3.1, and their values are tabulated in Table 1.

3.3 *Discussions*

Figure 2 shows two MR images Test_Image#A and Test_Image#B and their respective enhanced output images. The original Test_Image#A image consists of two main

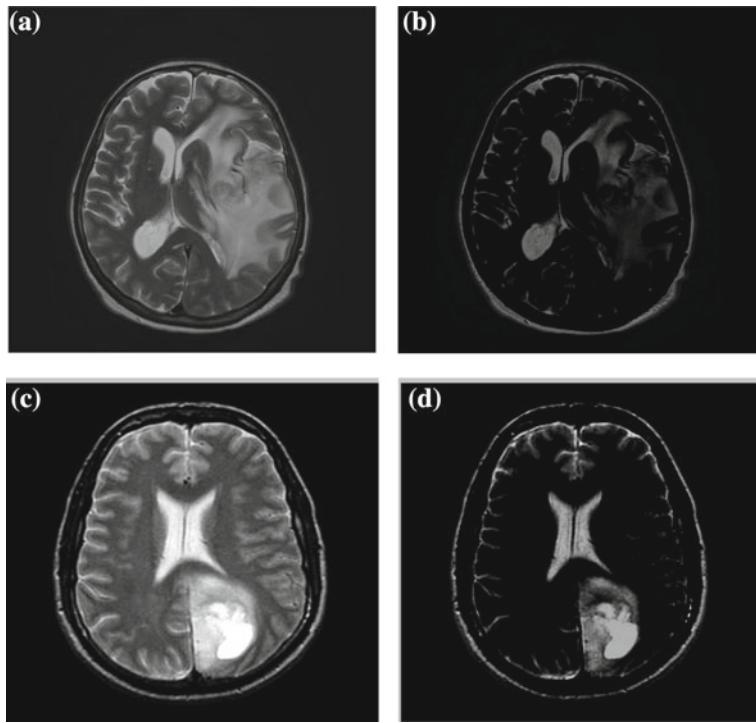


Fig. 2 **a** Original MRI image Test_Image#A. **b** Enhanced Test_Image#A. **c** Original Test_Image#B. **d** Enhanced Test_Image#B

Table 1 Values of CII and PSNR for proposed enhancement method

Images	CII	PSNR	
		Original	Enhanced
Test_Image#A (MR-T1)	2.2923	2.6917	5.1413
Test_Image#B (MR T2-PD)	1.4367	2.5509	4.2925
Test_Image#C (MR-T2-PD)	1.3793	3.2965	6.2461
Test_Image#D (MR-T1)	1.5818	3.5772	6.9678

structures that are the central and the lower oval shape structure on the left half of the image. These structures are visible very distinctly in the enhanced image. Their background is largely suppressed, and the large patch on the right-hand side which is not very significant is omitted to some extent. In Test_Image#B, the original image has two high-intensity regions on the centre and the lower right half of the image which are merged with the moderate-intensity background. After enhancement, the regions are more accentuated by curtailing the background features. Table 1 shows good performance metrics with incrementing values of CII and PSNR.

4 Conclusion

A highly rigorous study for brain MR images is required for proper detection and effective treatment of brain tumor. In this paper, an improved approach for MR image enhancement is suggested which uses morphological filtering. Morphological filters are one of the simplest filters in terms of their mathematical complexity and implementation. The proposed filter which uses a disk-shaped structuring element of size 35 along with top-hat and bottom-hat operators is considered. The results produced ensure a significant improvement in contrast depicted by the incrementing values of the quality assessment metrics, i.e., CII and PSNR. The observed values of CII are comparable to the values obtained from the existing methods. Since in this enhancement technique only significant features are preserved, the method may further assist and simplify the segmentation and classification stages.

References

1. Types of Brain Tumors. <https://www.abta.org>
2. Alankrita, A. R., Shrivastava, A., Bhateja, V.: Contrast improvement of cerebral mri features using combination of non-linear enhancement operator and morphological filter. In: Proceedings of IEEE International Conference on Network and Computational Intelligence (ICNCI), Zhengzhou, China, vol. 4, pp. 182–187 (2011)
3. Bahadure, N.B., Ray, A.K., Thethi, H.P.: Comparative approach of MRI-based brain tumor segmentation and classification using genetic algorithm. *Int. J. Digital Imag.* 1–13 (2018)
4. Somasundaram, K., Kalaiselvi, T.: Automatic brain extraction methods for T1 magnetic resonance images using region labeling and morphological operations. *Comput. Biol. Med.* **41**(8), 716–725 (2011)
5. MRI sequences (overview), <https://radiopaedia.org/articles/mri-sequences-overview>
6. Akram, M.U., Usman, A.: Computer aided system for brain tumor detection and segmentation. In: IEEE International Conference on Computer Networks and Information Technology (ICCNIT), pp. 299–302 (2011)
7. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **9**(5), 889–896 (2000)
8. Lidong, H., Wei, Z., Jun, W., Zebin, S.: Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement. *IET Image Proc.* **9**(10), 908–915 (2015)
9. Magudeeswaran, V., Singh, J.F.: Contrast limited fuzzy adaptive histogram equalization for enhancement of brain images. *Int. J. Imaging Syst. Technol.* **27**(1), 98–103 (2017)
10. Kharrat, A., Benamrane, N., Messaoud, M. B., Abid, M.: Detection of brain tumor in medical images. In: 3rd IEEE International Conference on Signals, Circuits and Systems (SCS), pp. 1–6 (2009)
11. Benson, C.C., Lajish, V.L.: Morphology based enhancement and skull stripping of MRI brain images. In: IEEE International Conference on Intelligent Computing Applications (ICICA), pp. 254–257 (2014)
12. Hassanpour, H., Samadiani, N., Salehi, S.M.: Using morphological transforms to enhance the contrast of medical images. *Egypt. J. Radiol. Nucl. Med.* **46**(2), 481–489 (2015)
13. Verma, R., Mehrotra, R., Bhateja, V.: A new morphological filtering algorithm for pre-processing of electrocardiographic signals. In: Proceedings of the Fourth International Conference on Signal and Image Processing (ICSIP), pp. 193–201. Springer, India (2013)

14. Tiwari, D. K., Bhateja, V., Anand, D., Srivastava, A., Omar, Z.: Combination of EEMD and morphological filtering for baseline wander correction in EMG signals. In: Proceedings of 2nd International Conference on Micro-Electronics, Electromagnetics and Telecommunications, pp. 365–373. Springer, Singapore (2018)
15. Raj, A., Srivastava, A., Bhateja, V.: Computer aided detection of brain tumor in magnetic resonance images. *Int. J. Eng. Technol.* **3**(5), 523–532 (2011)
16. Bhateja, V., Urooj, S., Mehrotra, R., Verma, R., Lay-Ekuakille, A., Verma, V.D.: A composite wavelets and morphology approach for ECG noise filtering. In: International Conference on Pattern Recognition and Machine Intelligence, pp. 361–366. Springer, Berlin (2013)
17. Bhateja, V., Devi, S.: A novel framework for edge detection of microcalcifications using a non-linear enhancement operator and morphological filter. In: IEEE 3rd International Conference on Electronics Computer Technology (ICECT), vol. 5, pp. 419–424 (2011)
18. Bhateja, V., Urooj, S., Verma, R., Mehrotra, R.: A novel approach for suppression of powerline interference and impulse noise in ECG signals. In: IEEE International Conference on Multi-media, Signal Processing and Communication Technologies, pp. 103–107 (2013)
19. Chaddad, A., Tanougast, C.: Quantitative evaluation of robust skull stripping and tumor detection applied to axial MR images. *Brain Inform.* **3**(1), 53–61 (2016)
20. Verma, R., Mehrotra, R., Bhateja, V.: An integration of improved median and morphological filtering techniques for electrocardiogram signal processing. In: IEEE 3rd International Conference Advance Computing, pp. 1223–1228 (2013)
21. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, pp. 689–794. Pearson Education (2009)
22. Srivastava, A., Raj, A., Bhateja, V.: Combination of wavelet transform and morphological filtering for enhancement of magnetic resonance images. In: Digital Information Processing and Communications, pp. 460–474. Springer, Berlin (2011)
23. Jain, A., Singh, S., Bhateja, V.: A robust approach for denoising and enhancement of mammographic images contaminated with high density impulse noise. *Int. J. Converge. Comput.* **1**(1), 38–49 (2013)
24. The Whole Brain Atlas, <http://www.med.harvard.edu/aanlib/home.html>
25. The Internet Brain Segmentation Repository, <https://www.nitrc.org/projects/ibsr/>

Author Index

A

- Ali, Noura, 415
Alva, Sathyavathi R., 453
Ansari, Nazneen, 363
Antony, P. J., 453
Arrmuthan, M., 405
Arun, Amaithi Rajan, 151
Arun Kumar, S., 53
Arya, Anu, 833
Arya, Meenakshi S., 823
Aski, Vidyadhar J., 93
Auty, Smita, 121, 315
Ayare, Prachi, 659

B

- Babu, S. Suresh, 83
Bairathi, Divya, 111
Bari, Santosh G., 121, 315
Basak, Rohini, 423
Benita, B., 299
Bhadauria, Anuj Singh, 833
Bhagwat, Alok, 345
Bhalchandra, Parag, 171, 239
Bhateja, Vikrant, 833
Bhat, Ghulam Mohiuddin, 161
Bhavani, N. P. G., 129
Bhogal, Rosepreet Kaur, 799
Bhoyer, Pravin Kumar, 71
Bhukya, Raju, 615
Bobhate, Ramchandra, 659
Bose, Rajesh, 559, 603

C

- Cadersaib, Z., 333
Chakraborty, Dhriti, 491

- Chattopadhyay, Matangini, 491
Chaudhary, Neera, 103
Chavan, Mahesh, 327
Chavan, Shamkumar, 327
Cheerkoot-Jalim, Sudha, 333, 547
Cherian, Anish, 83
Chitrao, Pradnya Vishwas, 71, 345
Chooramun, N., 333

D

- Dambe, Swapnil, 249
Darbari, Manuj, 729
Datta, Arnab, 399
Dave, Meenu, 789
Deep, Bhavya, 749
Deepshikha, 501
Deshmukh, Nilesh, 239
Deshmukh, Rutuja, 121, 315
Deshpande, Aditi P., 205
Deshpande, Meena, 571
Dey, Raktim Kumar, 559, 603
Dhal, Sunil, 695
Dhareshwar, Shweta, 759
Divekar, (Brig) Rajiv, 71

E

- Elaheebocus, S., 333

G

- Ganesan, A., 129
Garg, Saloni, 23, 43
Geethanjali, B., 271
Geetha, R., 229, 289
George, Shine K., 463
Gobin-Rahimbux, B., 333

Gondkar, Raju Ramakrishna, 83
 Gooda Sahib-Kaudeer, N., 333
H
 Gopalani, Dinesh, 111
 Gopalan, Santhosh Kumar, 463
 Gopal, Kanishka, 547
 Gorewar, Harish V., 217
 Gottala, Vineel Abhinav, 431
 Guchhait, Prasun, 93
 Gulati, Savy, 799

J
 Hafiz, Abdul Mueed, 161
 Hambarde, Kailas, 171
 Heenaye-Mamode Khan, M., 333
 Husen, Shaikh, 171

K
 Jagathy Raj, V. P., 463
 Jain, Anshul, 259
 Jain, Sushma, 739
 Jaiswal, Ajay A., 217
 Janaki, E., 651
 Janani, P., 195
 Jawandhiya, Pradip M., 13
 Joshi, Nikhilesh, 521
 Joshi, Nisheeth, 749
 Jovanovic, Raka, 305

K
 Kamila, Nilayam Kumar, 695
 Karanatakam, Dharani, 299
 Karthika, S., 229, 289
 Karthikeyan, V., 129
 Katti, Vijayalakshmi S., 759
 Kaur, Pardeep, 539
 Khamitkar, Santosh, 171
 Khara, Satvik, 279
 Kishnah, S., 333
 Kochhar, Aarti, 539
 Komathy, K., 443
 Kothari, Komal, 279
 Krishna, Jaya, 705
 Kulkarni, Aditya, 249
 Kulkarni, Govind, 171, 239
 Kumar, Arun, 615
 Kumar, Arvind, 473
 Kumar, Ashish, 1
 Kumar, Ashwani, 513, 595
 Kumar, Indeewar, 1
 Kumar, Manoj, 183
 Kumar, Mohit, 513
 Kumar, Raj, 473
 Kumar, Ravindra, 23, 43
 Kumar, Suresh, 103

Kumbhare, Vijay Rao, 481

L
 Laddha, Shilpa, 13
 Lalitha, D., 669, 675
 Lobo, Vivian Brian, 363
 Lohi, Shantanu Anandrao, 217
 Lohi, Snehal A., 217

M
 Mahender, C. N., 205
 Majumder, Manoj Kumar, 481
 Malhotra, Jyoti, 659
 Mandot, Manju, 683
 Manjare, Ashwini, 823
 Mary Metilda, M. I., 675
 Mathur, Iti, 749
 Mathur, Rajeev, 33
 Mavani, Umang, 363
 Mishra, Rupesh, 363
 Mohamed Ismail, A., 651
 Muley, Aniket, 239
 Muralidhara, B. L., 271

N
 Nageswara Rao, J., 625
 Naik, Payal, 823
 Nair, Priyanka, 705
 Narwekar, Krupa, 531
 Natarajan, S., 195
 Nidhi, Ritu, 383
 Nigam, Mansi, 833

P
 Pal, Bijoy, 399
 Paltani, Punya Prasanna, 481
 Pandi, Gayatri S., 715
 Patel, Chhayaben Nitin, 373
 Patel, Vishva Nitin, 373
 Pednekar, Aditi, 363
 Pereira, Naomi Christianne, 363
 Ponnusamy, R., 405
 Poulraju, Benita, 299
 Pradhan, Prabodh Kumar, 695
 Pragatheevaran, Sridharan, 151
 Preeti, 539
 Pushpalatha, N., 779

R
 Ramesh, G. V., 633
 Randhawa, Sukchandan, 739
 Rathod, Akshada, 659
 Rathore, Hemant, 139
 Ray, Debasish, 633

Roy, Sandip, 391, 399, 559, 603
Roy, Uttam K., 423

S

Sachan, Madhurendra, 729
Sachdeo, Rajneeshkaur, 659
Saha, Amit, 391
Sahay, Sanjay K., 139
Sai Krishna, P., 129
Saini, Dinesh Kumar, 643
Saini, Monika, 1
Sai Satyanarayana Reddy, S., 779
Samsani, Surekha, 431
Sandhiyaa, B. Y., 643
Sarddar, Debabrata, 559, 603
Sarkar, Indranil, 399
Sarode, Sambhaji, 659
Sarode, Tanuja, 521
Satya Narayana, K., 633
Saxena, Archana B., 789
Seal, Dibyendu, 423
Sehgal, Pranav, 93
Shafiya, Shaik, 129
Shah, Aagam, 279
Shah, V. A., 531
Sharma, Ashu, 139
Sharma, Banabari lal, 615
Sharma, Jitendra Kumar, 615
Sheikh, Aysha Munir, 33
Sheikh, Kamrul Haq, 823
Shukla, Piyush, 585
Singh, Pawan, 513
Singh, Tanya, 259, 383
Sinha, Aditya Kumar, 195
Sowmya, K., 759
Sridevi Sathya Priya, S., 299
Srivastava, Devesh Kumar, 705
Srividhya, V., 129
S, Sushitha, 759

Subhash Chandra, N., 779
Sujatha, K., 129
Suthanthira Devi, P., 229, 289

T

Taduri, Srinath, 771
Tamsekar, Preetam, 171
Tarrad, I. F., 415
Thakkar, Umang, 279
Thakur, Manav Ashok, 121, 315
Thirunavukkarasu, Varatharajan, 151
Tiwari, Dimple, 183
Tiwari, Preeti, 585
Tuba, Eva, 305
Tuba, Milan, 305

V

Vaishnav, Arun, 683
Vandana, B. S., 453
Veena, M. B., 571
Veeraiah, D., 625
Vengata Krishnan, K., 669
Venkatesulu, M., 53
Venugopal, P., 289
Verma, Siddhant, 195
Vetriselvi, V., 151
Vijaya Prakash, Rajanala, 771
Vijayaragavan, P., 405

W

Wandra, K. H., 715
Wasi, Sheeban, 729

Y

Yadav, Subodh, 615
Yammiyavar, Pradeep, 501
Youssef, M. I., 415