



申请代码	G02
受理部门	
收件日期	
受理编号	7149720033



国家自然科学基金 申 请 书

(2014 版)

资助类别：	重大项目		
亚类说明：	课题申请		
附注说明：	大数据环境下的商务管理		
项目名称：	面向大数据的商务分析与计算方法以及支撑平台研究		
申 请 人：	刘业政	电 话：	055162904991
依托单位：	合肥工业大学		
通讯地址：	合肥市屯溪路192号12#信箱		
邮政编码：	230009	单位电话：	0551-62901115
电子邮箱：	liuyezheng@hfut.edu.cn		
申报日期：	2014年08月25日		

国家自然科学基金委员会



基本信息

申请人信息	姓名	刘业政	性别	男	出生年月	1965年09月	民族	汉族
	学位	博士	职称	教授	每年工作时间（月）		6	
	电话	055162904991		电子邮箱	liuyezheng@hfut.edu.cn			
	传真			国别或地区	中国			
	个人通讯地址	合肥市屯溪路192号12#信箱						
	工作单位	合肥工业大学/管理学院电子商务研究所						
	主要研究领域	电子商务, 商务智能, 数据挖掘, 个性化推荐系统, 智能决策理论与方法, 群决策理论与方法						
依托单位信息	名称	合肥工业大学						
	联系人	岳峰	电子邮箱	yuefeng@hfut.edu.cn				
	电话	0551-62901115	网站地址	www.hfut.edu.cn				
合作研究单位信息	单位名称							
	西南交通大学							
项目基本信息	项目名称	面向大数据的商务分析与计算方法以及支撑平台研究						
	英文名称	Research on Business Analytics, Computing Techniques and Platforms in the Context of Big Data						
	资助类别	重大项目			亚类说明	课题申请		
	附注说明	大数据环境下的商务管理						
	申请代码	G02						
	基地类别	“过程优化与智能决策”教育部重点实验室						
	研究期限	2015年01月 -- 2019年12月						
	申请经费	325.0000万元						
中文关键词	商务管理；大数据分析；大数据支撑平台；商业情报分析；个性化推荐							
英文关键词	Business Management；Big Data Analytics & Computing；Big Data Analytics Platform；Business Intelligence & Analytics；Personalized Recommendation							



中文摘要	<p>本课题以大数据环境下商务管理的新需求为指引，按照“可用信息—可用知识—知识服务—案例示范”的思路，系统研究面向商务管理的大数据分析和计算的核心方法以及关键技术，构建面向商务管理的内外部数据获取系统、计算实验平台以及原型系统。具体研究内容有：（1）面向商务管理的大数据获取方法与可视化技术；（2）面向商务管理的大数据分析与计算方法；（3）基于网络大数据的商业情报分析方法；（4）大数据驱动的个性化用户建模与推荐方法；（5）面向商务管理大数据支撑平台的核心理论与关键技术；（6）案例研究：基于大数据的汽车品牌管理系统。以上研究对于解决面向商务管理的大数据可用性评估、超高维稀疏性大数据的快速分析、中文主题识别、个性化推荐的多目标优化以及大数据支撑平台的体系结构设计等关键科学难题，构建面向商务管理的大数据分析与计算理论、方法和技术体系，促进大数据环境下我国商务管理理论与实践水平的发展具有重要意义。</p>
英文摘要	<p>Guided by the new demands of business management in the context of Big Data and the thought of “Available Information - Available Knowledge - Knowledge Service - Application Demonstration”, this project will investigate systematically the big data analytics methods and computing techniques for business management, and build the data acquisition system, computational experiment platform and prototype system for evaluation and demonstration. This project will conduct researches from the following six aspects. 1) Big Data acquisition methods and visualization techniques for business management; 2) Big Data analytics and computing methods for business management; 3) Business intelligence and analytics based on internet data; 4) Personalized user modeling and recommendation methods based on big data; 5) Core theories and key techniques of the Big Data analytics platform for business management; 6) Case study: an automobile brand management system in the context of Big Data. The research focuses on addressing the key scientific issues such as usability evaluation of Big Data, rapid analysis of Big Data with ultra-high dimension and sparsity, Chinese topic identification, multi-objective optimization of personalized recommendation and architecture design of the Big Data analytics platform for business management. It will be of great significance to leverage our research to enrich the theories, methods and techniques of Big Data analytics and computing for business management, and promote the development of business management theory and practice in the context of Big Data.</p>



项目组主要参与者（注：项目组主要参与者不包括项目申请人）

编号	姓名	出生年月	性别	职 称	学 位	单位名称	电话	电子邮箱	项目分工	每年工作时间（月）
1	杨善林	1948-10-05	男	教授	硕士	合肥工业大学	13905696729	hgdysl@gmail.com	总体规划与研究应用方案设计	3
2	倪志伟	1963-11-05	男	教授	博士	合肥工业大学	13966051682	nzwgd@hfut.edu.cn	大数据分析 with 计算方法	5
3	李军	1967-05-14	女	教授	博士	西南交通大学	13708080220	lijun@home.swjtu.edu.cn	大数据获取与评估	5
4	何建民	1961-03-05	男	教授	博士	合肥工业大学	13956089508	hejianmin@hfut.edu.cn	商务分析方法	5
5	姜元春	1980-09-28	男	副研究员	博士	合肥工业大学	13856955921	ycjiang@hfut.edu.cn	个性化用户建模与推荐方法	6
6	徐进	1978-02-14	男	副教授	硕士	西南交通大学	13980881921	jinxu@home.swjtu.edu.cn	大数据可视化技术	5
7	丁帅	1984-05-24	男	副研究员	博士	合肥工业大学	13855169247	dingshuai@hfut.edu.cn	大数据支撑平台	6
8	王晓佳	1982-02-18	男	讲师	博士	合肥工业大学	13866720415	tonysun800@sina.com	大数据分析 with 计算方法	6
9	孙见山	1987-03-01	男	讲师	博士	合肥工业大学	18956561661	sunjs9413@hfut.edu.cn	商业情报分析方法	6

总人数	高级	中级	初级	博士后	博士生	硕士生
36	8	2			20	6



经费申请表

(金额单位: 万元)

科目	申请经费	备注 (计算依据与说明)
一. 研究经费	243.7500	
1. 科研业务费	170.0000	
(1) 测试/计算/分析费	30.0000	数据处理、仿真及优化计算。
(2) 能源/动力费	35.0000	水、电等能源动力费用 (实验用云计算平台的能源消耗)。
(3) 会议费/差旅费	45.0000	组织或参加相关学术会议、调研费用、差旅费用。
(4) 出版物/文献/信息传播费	60.0000	国内外专业数据检索、入网、资料购买, 调研报告及学术论著出版、专利申请等费用。
(5) 其他	0	
2. 实验材料费	30.0000	
(1) 原材料/试剂/药品购置费	30.0000	购买数据使用权、复印机耗材、打印机耗材等费用。
(2) 其他	0	
3. 仪器设备费	33.7500	
(1) 购置	33.7500	添置搭建云计算实验平台所必须的软硬件; 数据分析软件; 系统仿真软件。
(2) 试制	0	
4. 实验室改装费	10.0000	为改善资助项目研究的实验条件, 对实验室进行改装所开支的费用。
5. 协作费	0	
二. 国际合作与交流费	32.5000	
1. 项目组成员出国合作交流	20.0000	前往国外开展与本项目相关的短期合作研究、国际学术会议。
2. 境外专家来华合作交流	12.5000	邀请国外专家访问交流。
三. 劳务费	32.5000	直接参加项目研究的研究生、博士后人员的劳务费用。
四. 管理费	16.2500	项目依托单位为组织和支持本资助项目研究而支出的费用, 不得超过申请经费的5%。
合计	325.0000	
与本项目相关的其他经费来源	国家其他计划资助经费	0
	其他经费资助 (含部门匹配)	0
	其他经费来源合计	0



报告正文

参照以下提纲撰写，要求分栏目撰写，条目清晰，标题突出。

（一）立论依据与研究内容

0、课题说明

本研究课题为国家自然科学基金委员会重大项目申请“大数据环境下的商务管理研究”之所属课题五。该重大项目由清华大学、复旦大学、西南交通大学、合肥工业大学、哈尔滨工业大学联合申请，依托单位为清华大学。项目申请人陈剑，主要成员包括：杨善林、黄丽华、贾建民、陈国青、刘业政、叶强、徐心、刘红岩、戴悦等。其所属课题设置如下：

课题一：大数据环境下的商务行为机理研究

课题二：大数据环境下的顾客洞察与市场营销策略研究

课题三：大数据环境下的运营策略优化与协调研究

课题四：基于大数据的商务智能与模式创新研究

课题五：面向大数据的商务分析与计算方法以及支撑平台研究

本研究课题与该重大研究项目其他所属课题的关系如图0所示。

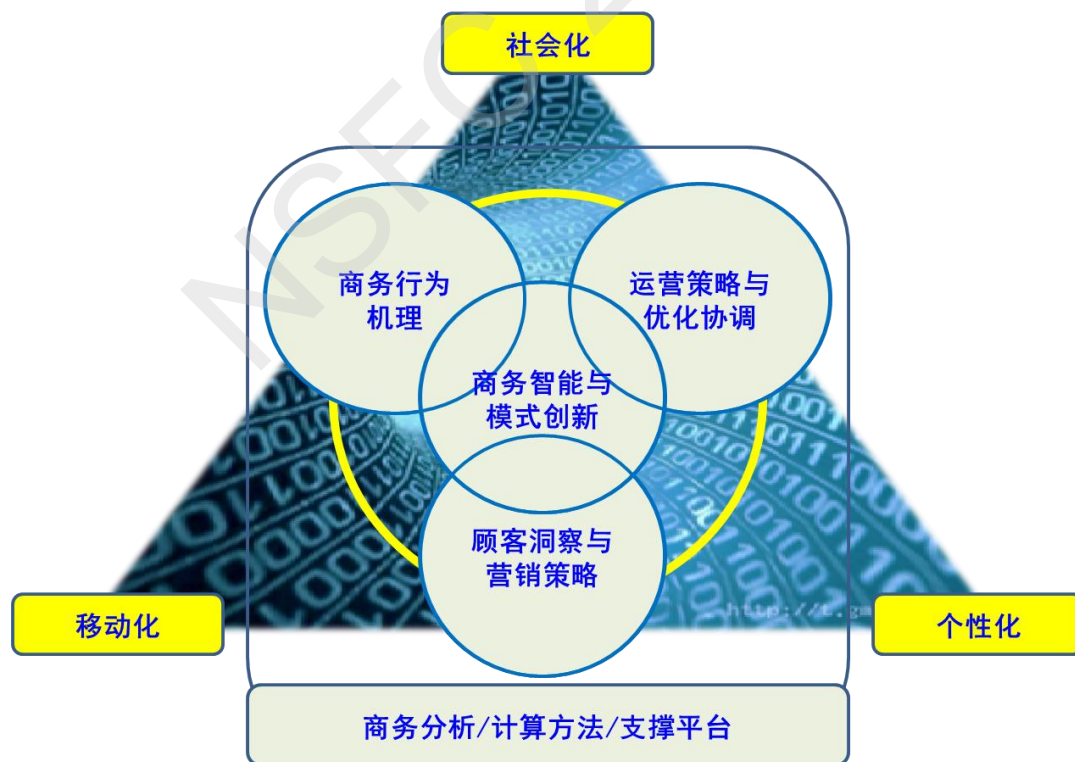


图0 “大数据环境下的商务管理研究”重大项目



1、课题研究意义、国内外研究现状及分析。需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录。

1.1 立项背景与研究意义

大数据与自然资源、人力资源一样是国家、组织的重要战略资源，是推动国民经济、社会和科学技术发展的巨大动力。大数据分析与应用已经成为世界各国构建创新发展模式和提升国家长期竞争力的新的战略领域，正改变着网络安全、情报分析、危机响应、市场营销的方式，迅速成为商务管理、金融服务、社会治理、计算机科学、国家安全等领域的前沿热点问题。近年来，众多的数据科学家、企业和政府部门开始投入到大数据相关理论与技术的研究进程中，在大数据的产生与集聚、组织与管理、分析与发现、应用与服务等多个方面获得了突破性进展，相关成果在商务管理中也得到了有效的应用。例如，亚马逊利用大数据挖掘技术预测客户的兴趣偏好和优化结果排序，提高营销精准度；百度通过分析游戏爱好者的搜索行为，向网络游戏运营商提供大数据分析支持；阿里巴巴则运用大数据技术成功预见金融危机的来临，有效管控了阿里巴巴及其客户的经营风险。

大数据作为企业的一种重要战略资源，突破面向商务管理的大数据分析与计算的核心方法和关键技术（本课题），发现隐匿于商务大数据中的有用知识，是实现大数据大价值的核心所在，也是深入研究大数据环境下人、组织与市场的交互行为规律（课题一）以及大数据驱动的企业营销管理变革（课题二）、运营管理变革（课题三）与商务决策模式创新（课题四）的基础。过去几十年来，一方面，数据库、搜索引擎、电子商务、社交网络等的研究与开发，在可扩展的以数据为中心的技术设计方面积累了丰富的经验，并产生了能较好地适合大数据分析的云计算以及其他并行、分布式平台；另一方面，在机器学习、数据挖掘、统计学以及算法理论方面的创新也产生了许多适合于处理大数据的数据分析方法。但由于大数据资源的分散、多样、高维、稀疏、不确定等复杂性，在面向商务管理的大数据分析、计算与有效应用方面还存在着一些重要的困难（详见国内外研究现状与发展动态分析），具体地：

1. 面向商务管理的大规模数据获取方面。大数据的商务应用更加倾向于跨平台多元化的数据应用，特别关注不同数据拥有者之间的数据聚合、共享与协同。支持商务管理的大数据资源以多种形态碎片化地散落在不同的数据源之中，数据往往存在着质量参差不齐、时效性要求过高且并不总能满足商务管理需求等问题；同时，大数据的非结构化和时空域特性给数据使用者也带来了理解上的困难。因此，发现并减轻数据获取瓶颈、快速评估数据的质量及其可用性、提出多源数据的过滤、融合及可视化展示方法、设计基于并行和分布式硬件平台的数据管理策略、开发简单且具有通用性的数据获取接口成为大数据分析与计算的重要挑战。



2. 面向商务管理的大数据分析与计算方面。大数据分析计算面临着许多挑战，其中大数据分析的计算方法和大数据分析的统计检验是两个最主要的挑战。**(1) 大数据分析的计算方法问题。**大数据分析面临着富媒体数据、时空域数据、流数据、图结构数据等带来的巨大挑战，例如，文本等富媒体数据分析中的情感挖掘、文档主题模型构建、语义分析与提取等问题，时空域数据分析中的时间连贯性和空间尺度的维持问题，商务流数据分析中的管理统计风险与概念漂移等问题，以及图结构数据中的海量结点和多模网络问题，均需要开展持续的、深入的研究才能逐步得到有效解决；**(2) 大数据分析的统计检验问题。**从数据到知识的推断必须能够通过统计检验才能保证所获取的知识的质量，但影响统计推断质量的因素有很多，如超高维、异构性、抽样偏差、未知的数据出处、极端稀疏的长尾数据等，虽然统计领域开发出的一些工具大体上能解决这些问题，但在大数据环境下，现有统计工具所基于的与数据集特征和抽样方式相关的假定可能不再有效，误差评估和统计诊断工具自身的计算过程可能是计算上不可行的。因此需要开发新的适用于大数据的统计推断与机器学习方法。

3. 面向商务管理的大数据分析系统的构建与应用方面。**(1) 分布式数据资源的有效利用问题。**大数据分析系统中的数据可能来自于不同的组织、分布存储在不同的物理系统中，如何发现并协同利用这些数据资源，实现商务管理目标，既存在技术上的挑战，也存在着管理上的挑战；**(2) 复杂商务分析任务的有效计算问题。**面对复杂的商务分析需求，在并行计算框架下，商务分析任务的分解、计算工作流的优化设计、各计算任务间的协同管理与集结、以及突发计算事件发生时的动态调度等问题，都需要开展系统深入的研究；**(3) 人的有效参与问题。**在大数据分析过程中，许多数据集需要进行语义理解，而这超出了目前计算方法的能力，因而需要人在整个分析过程中的深入参与，参与者（数据分析师、管理者、客户或客户众包等）通过对整个数据分析过程中的假设框架、权衡管理（精度与速度）、问题选择等的判断提供相应的输入，此时，有效的可视化和界面设计，特别是在连接人的判断力与数据分析算法方面，面临着许多挑战；**(4) 软件开发问题。**开发连接顶层的数据分析规范与底层分布式系统体系结构的软件组件，为终端用户开发易用的、易理解的、支持人工参与的大数据分析系统，以及软件开发所面对的诸如数据持续到达的流环境、分布式环境等，都面临着巨大的挑战。

基于上述分析，本课题将系统研究面向商务管理的多源大数据获取、评估与融合方法，研究面向商务管理的大数据系统建模和分析计算方法，研究面向商务管理大数据支撑平台的体系结构、并行计算架构下的数据管理理论和基于工作流的大数据分析系统优化设计理论、人机协同机制与可视化技术等关键科学问题，并从基于网络大数据的商业情报分析、大数据驱动的个性化信息推荐两个具体应用领域研究具有针对性的解决方案，开发面向商务管理的内外部数据获取系统、计算实验平台以及原型系统，形成一套较为完整的面向商务管理的大数据分析与计算理论、方法和技术体系，为在企业和组织的商务管理创新中高效使用大数据资源奠定坚实的理论方法和技术基础。



本课题的研究意义在于：

(1) 从面向商务管理的大数据获取、评估、融合、系统建模、分析与计算以及基于网络大数据的商业情报分析、大数据驱动的个性化信息推荐等方面，系统深入地研究面向商务管理的大数据分析与计算理论、方法及其关键技术，对进一步丰富现代商务管理的理论与方法体系具有重要的理论意义。

(2) 大数据环境下的商务管理具有广阔的应用前景，开发面向商务管理研究与应用的支撑平台不仅对促进大数据在商务管理中的应用，创新商务决策模式，优化营销管理与运营管理过程，提高营销管理与运营管理效率，增强企业核心竞争力具有重要的现实意义，而且对开展“大数据环境下的商务管理”重大项目的协同研究具有重要的支撑作用。

(3) 研究面向大数据的商务分析与计算方法以及支撑平台，涉及到云计算、大数据、信息系统技术和商务管理理论、行为科学、协同理论以及其他管理理论和方法，本课题的研究成果对促进相关学科的交叉融合与发展具有一定的推动作用。

1.2 国内外研究现状及发展动态分析

随着经济全球化进程的加速，以及大数据、云计算、物联网等新兴信息技术的快速普及，研究面向商务管理的大数据分析与计算的核心方法和关键技术，开发面向商务管理研究与应用的支撑平台已经成为跨国企业巨头争相追逐的一个战略制高点，也是支持“大数据环境下的商务管理”重大项目研究的基础性工作。根据“大数据环境下的商务管理”重大项目中课题五“面向大数据的商务分析与计算方法以及支撑平台”的项目指南，我们从大数据获取方法与可视化技术、面向商务管理的大数据分析与计算方法、大数据环境下商业情报分析方法、大数据驱动的个性化用户建模与推荐方法、面向商务管理大数据支撑平台的核心理论与关键技术、以及大数据在商务管理中的应用实践等方面综述相关的国内外研究现状。

(1) 关于大数据获取方法与可视化技术方面的研究

商务管理的复杂性，以及大数据生成环境的开放性和不确定性，使得支持商务管理的大数据资源以多种形态碎片化地散落在不同的数据源之中，跨平台多源化的数据往往存在着质量参差不齐、时效性要求过高且并不总能满足商务管理需求等问题，进而造成面向商务管理的大数据并不总是可用的且呈现出价值稀疏性特征；数据结构的多样性、时空域特性的客观差异，以及商务管理应用中的持续计算要求，使得在多源数据的融合、使用者对数据的理解方面都出现了很大的困难。因此，研究面向商务管理的多源异构大数据获取方法、可用性评价方法、数据融合方法和可视化技术已经成为大数据环境下商务管理创新亟需解决的首要问题，是实现面向商务管理的大数据分析的基础。具体研究进展综述如下。

在大数据获取研究方面。大数据环境下，数据分析已由业务驱动转变为数据驱动，数据获取已由传统的找、抓，变为目前的筛选、抽样^[1]，针对由海量的 Web 文档、在线社会网络及在线评论等非结构化数据和声音、视频等富媒体数据组成的内外部商业数据，对大数据的感知与获取工作（尤其是非结构化/半结构化数据的高效处理和表示、数据的去冗余^[2]、流数据获取过程中的采样方法和模式识别方法^[3]等）已经成为大数据研究的重要内容^[4]。



在针对商务管理的大数据获取研究方面，文献[5]通过实证指出在商务分析中 ETL (Extraction-Transformation-Loading) 是必须的，文献[6]援引文献[7]的成果将面向商务管理大数据分析过程分解为目标确定 (Domain understanding & KDD Goals)、数据选择 (Selection & Addition)、数据清洗、数据转换、数据挖掘、可视化和反馈等典型阶段，说明在商务管理中根据目标进行数据选择的重要性。Buelens 等 2014 年 7 月份在 Statistics Netherlands 上发布的一篇 Discussion Paper^[8]则从更广的视角指出大数据的可选择性 (Selectivity of Big data) 问题的研究意义。

在大数据可用性评估研究方面。有效地评估数据的可用性是数据选择和筛选的基本保障，数据可用性评估是一个层次分类的概念，每个类逐级分解成具体的评估度量，并通过对这些度量的计算与集成获得数据的整体质量，目前国内外相关研究工作主要聚焦于评估指标体系、评估模型和评估方法等。**对评估指标体系的研究**，文献[9]使用完整性、正确性、一致性、似然性和时效性等多个客观度量对数据质量评价进行了系统性分析；文献[10]回顾了与数据可用性以及数据质量相关的研究进展，详细介绍了数据质量的 16 个评估度量；文献[11]则深入讨论了数据质量评估度量的选择方法。**对评估模型的研究**，文献[12]在对信息质量度量维度归类的基础上，构建了一个信息质量层次评估模型，同时考虑了各质量维度指标的权重；文献[13]在数据质量标准 ISO/IEC25012 和 PDQM (Portal Data Quality Model) 模型的基础上提出了一种评估门户网站中的数据质量的模型。**对评估方法的研究**，文献[14]借鉴信息完整性评估的常用控制策略，提出了一种基于控制矩阵的数据质量综合评估方法；文献[15]还通过对许多较为成熟的数据质量评估方法和工具的有效整合，研究了普适性更强的综合数据质量评估方法学。

在大数据融合研究方面。文献[16]指出多学科多技术方法的数据集成与融合是大数据的四大挑战之一，文献[17]给出了不同数据融合方式的综述和分类。与大数据融合相关的国内外研究工作主要集中于多源异构数据的采集、传输和聚合策略设计，通过构建基于树^[18]或者是基于节点聚类^[19]的数据聚合协议，实现对多源异构数据的有效集成。文献[20]详细分析了基于无线传感器网络的数据聚合过程中存在的数据安全风险，建立了一种有利于促进数据安全提升的多源数据聚合协议框架；文献[21]构建了面向非相关性数据的传感器网络聚合函数，通过在数据融合节点对相似数据的冗余删减处理，以降低多源数据集成过程中的数据传输量；文献[22]分析了无线传感器网络中多源协同控制的数据集成问题，建立了一种考虑空间相关性的多源数据聚合方法；文献[23]研究了关注评估时间和空间相关性的自适应 MAC 协议，该协议的实施在提升数据聚合效率的同时还能显著减少传感器网络中的数据传输量，从而降低网络的整体能耗和延迟；文献[24]还提出了一种基于空间相关性的事件驱动型 MAC 协议。富媒体数据的融合一般采用社会标注的方法，以社会标签来表征流媒体的语义信息，文献[25]的研究结果表明无论是自动生成的元数据，社会标签都能够有效表达视频内容等富媒体的语义。文献[26]组合视觉特征和文本信息进行网页图像概念化，提出了 FMD 网页图像注释的新方法。文献[27]提出了一种语义视频注释的创新性方法，此方法通



过视觉特征、语音特征和频繁语义模型的集成挖掘完成语义视频注释。

在大数据的可视化技术研究方面。目前国内外相关研究工作主要聚焦于研究多维数据可视化、时空信息可视化和可视化在各特定领域的应用等方面。**对多维数据可视化研究**，平行坐标法^[28-30]广泛应用于高维和多变量数据的可视化，但是此方法应用于大数据，效果被边缘杂波减弱^[31]。文献[32]提出了一个新的框架，减少了边缘杂波的影响，提高了可视化聚类的效果。**对时空信息可视化研究**，文献[33]对已有的时空数据可视化技术所适用的时空数据类型和任务类型进行了综述，文献[34]提出了自定义计算方法保证大数据的实时可视化分析。文献[35]展示了一个名叫混合实体环境的可视化空间实现大量异构数据的可扩展性可视化。文献[36]提出了一个模糊时空信息表示模型，文献[37]在做时空数据可视化时使用树图做变量选择。**对商务信息可视化研究**，文献[38]介绍了商务信息可视化的工具。国内学者在信息可视化领域也取得了一些有影响力的成果^[39,40]。

现状分析：大数据生成环境的开放性和不确定性等特征给面向商务管理的大数据获取与可用性评估、大数据融合、大数据可视化等方面的研究带来了巨大挑战。通过对国内外研究现状的总结与分析可以看出，在数据获取与可用性评估、多源数据融合和数据可视化技术等方面已经取得了一些进展，但依然还存在着：①已有关于大数据的可用性评估模型或方法多是从数据源本身的质量特性去分析数据可用性，忽视了商务管理需求和数据间关联的重要影响，也未能考虑到大数据应用的即时性和动态性等特征。②对富媒体和流数据融合问题的研究已经开始关注多源数据的异构、相关性以及安全性等问题，但依然缺少面向商务管理的大数据语义集成方法和多维度融合框架。③在大数据可视化技术方面已有很多有影响力的成果，但仍然缺少有效的针对超高维数据和时间序列数据的可视化方法，缺少针对商务管理领域的时空信息可视化方法，缺少针对流数据的可视化自动生成技术。

(2) 关于面向商务管理的大数据分析与计算方法方面的研究

发现隐匿于大数据之中的有用知识是开展大数据研究的一项重要任务和价值所在，是实现大数据环境下商务管理的根本保证。然而，大数据环境下数据规模的急剧膨胀、数据形式的异常丰富以及数据存储的碎片化和分布式特征，使得传统的数据系统建模理论在应对大数据系统建模问题时显得极为乏力；同时，大数据不仅在数据规模、存储形式和高时效性需求上限制了传统数据分析方法的有效应用，也对具有较强学习能力的人工智能和数据挖掘方法形成了巨大挑战。为了在大数据发展浪潮中提升组织的知识获取能力和服务创新能力，越来越多的专家学者和 IT 企业开始投入到大数据系统建模以及大数据分析与计算方法的研究进程中，具体研究进展综述如下。

在大数据系统建模研究方面。大数据环境下，大数据的多源性、海量性、异构性以及由此产生的超高维、稀疏、非对称等数据特征使得在进行数据挖掘之前需要对大数据进行系统建模。大数据环境下系统模型的构建是对数据分析与计算过程中数据之间的约束、协作、交互等关联关系的系统化描述。**对数据语义层模型构建的研究。**文献[41]提出了基于文本片段的语义相似性度量方法，利用模式提取算法和模式聚类算法分析词和文本之间的多



种语义关系。文献[42]用名词结构对本体建模语言 OWL2 进行描述逻辑扩展,并将 SWRL 和 RIF 语言整合到 OWL2 中,使其可以用自身语法描述任意的 DL-safe 规则。**对数据表示层模型构建的研究。**文献[43]从时间、空间、关系等多个不同维度介绍了大数据的复杂性特征,指出复杂性是开展大数据研究必须关注的重点。文献[44]提出了一种基于数据库的本体模型,该模型允许用户基于应用领域对本体的概念进行汇总或修改。文献[45]提出了基于无监督方法的领域关系 (Feature-Of, Action-On, Type-of 和 Synonyms) 识别方法,该方法可以基于语料库分析领域术语,进而构建面向特定领域的浅层本体。文献[46]提出了一个允许资源自动验证和处理的一般框架,利用语义 Web 技术对语言学上的注释资源进行形式化描述,该方法可以对半结构化、非结构化数据进行有效描述。文献[47]分析了数据的商立方体 (quotient cube, QC) 语义描述方法对存储空间占用过大的问题,提出了使用 QC-Tree 结构的流数据立方体描述方法,该方法对于压缩数据立方的存储结构和节省内存开销效果显著。**对数据分析层面模型构建的研究。**文献[48]提出了一种适用于大规模流数据挖掘的计算模型 MUD。文献[49]研究了基于多核处理器的系统模型,该模型包含了数据集、数据分割、Map 处理、Reduce 处理和数据融合等基本要素。文献[50]将 IBM InfoSphere MDM 进行非结构化文本相关性 (EUTC) 扩展,从而将在非结构化数据中提取的信息集成到主数据集成系统 MDM 中。

在面向存储数据的知识发现研究方面。面向存储数据的知识发现已经取得了较多的研究成果,下面将从分类、聚类和关联关系发现三个方面进行现状回顾。**对分类查询的研究,**文献[51]研究了基于多层神经网络的超高维数据约简方法,该方法将高维数据转变为低维编码,提高了超高维数据的分析效率。文献[52]使用随机亚梯度下降算法 pegasos 近似最小化 SVM 的目标函数,所提算法在处理大型文本分类问题时具有很快的收敛速度;文献[53]使用逆向映射的方法化简集合,从而有效地训练支持向量机,得到基于最小二乘法的支持向量机分类器;文献[54]针对半结构化数据,利用改进的决策树算法在查询结果上自动构建一个带标签的分层分类树,具有较低的搜索代价和较好的分类效果;文献[55]设计了基于多目标遗传规划 (MOGP) 框架的非平衡数据分类模型,在少数类和多数类上均具有较好的分类精度。**对聚类研究,**文献[56]研究了面向大规模数据的谱聚类方法,所提方法用只保留最近邻的矩阵稀疏化方法以及 Nystrom 方法设计相似矩阵的近似方法,并在分布式框架下对稀疏化方法进行并行计算。文献[57]将基于随机游走的局部算法应用于大规模图聚类,使用集群的内部连通性作为聚类标准,理论证明所提方法在高连通图中具有很好的聚类效果。文献[58]使用随机游走的方法选择初始节点,以 Nibble 算法作为局部聚类算法,在近线性时间内完成大规模的图聚类。文献[59]在 SSC 算法 (Sparse Subspace Clustering) 的基础上提出了 SSSC 方法 (Scalable Sparse Subspace Clustering),在原数据集中抽取样本进行 SSC 聚类,将未抽取样本编码并分入已聚集的类中,解决了稀疏子空间聚类方法的可扩展性问题。文献[60]提出了基于密度峰值 (Density peaks) 的聚类方法,该方法可以有效克服异常数据的影响,提高聚类的效率;文献[61]根据高维数据集的自然特性将属性集分割成不同的属性组,提出



了考虑子属性空间权重的改进聚类算法 FG-k-means，该算法对高维空间中的噪音和缺失值具有很好的鲁棒性。《美国科学院院报》(PNAS)近年来出版了一系列数据聚类研究成果，对大数据环境下的聚类分析进行了深入探索^[62-64]。**对关联关系发现的研究**，文献[65]使用最大信息系数(MIC)来表示两个变量之间的依赖关系，提出了基于 MIC 的无参数统计方法来确定和区分关联关系。文献[66]提出 CoGAR 框架来支持受约束的广义关联规则挖掘，该框架利用多重分类标准防止非频繁项被过滤，用模式约束和机会主义信赖约束对显著的和冗余的规则进行区分。文献[67]研究了面向异构语义数据仓库的关联规则挖掘方法，该方法将本体编码获得的模式层知识反馈给传统的关联规则算法，提高关联规则挖掘的效果。

在面向流数据的知识发现研究方面。由于存储数据的知识发现方法无法适应流数据的连续性、多变性、快速性、无限性等特点，容易导致大量有用信息的丢失，因此，需要有针对性地研究面向流数据的知识发现方法。**对流数据抽样的研究**，文献[68]提出了一种新的抽样模型 FPS，保证生成的样本具有高质量且可以扩展到无限流中；文献[69]研究了利用双时间戳进行偏倚抽样的方法，通过所设计的衰减函数确定数据流元组的权重，根据该权重进行抽样，并利用平衡二叉树存储滑动窗口内数据流元组的权重，以提高数据流操作的响应时间；文献[70]将单一流连续抽样问题扩展到多重分布式数据流，从而获得全数据流样本；文献[71]提出了基于移动窗口的流数据抽样问题，用基于序列的窗口和基于时间戳的窗口进行链抽样和优先抽样。**对流数据挖掘的研究**，文献[72]利用加权全体分类器来挖掘概念漂移数据流；文献[73]开发了一种压缩模式流树(CPS-Tree)来捕获最新的流数据内容，同时还给出了一种自动树调整的概念来产生一个高度压缩的频率下降树结构；文献[74]提出了一种挖掘数据流任意大小滑动时间窗口内频繁模式的方法，可以在单遍扫描流数据的条件下及时捕获数据流上最新的模式信息；文献[75]设计了一种数据标记算法来决定当前滑动窗口内每个对象最合适的聚类标签，分析不同时间戳之间的聚类结果的关系；文献[76]提出 Learn++.NSE 和 Learn++.CDS 两种基于集合的非平衡数据概念漂移学习方法。文献[77]提出了一种不依赖于先验知识的时间序列分析方法，该方法可以对随机设定的非线性过滤问题进行有效分析。

现状分析：存储数据和流数据是大数据环境下两个主要的数据存在形式。针对存储数据和流数据的数据分析与计算是国内外学者研究的重点，已经形成了一批有较高影响力的理论成果，但在以下方面有待进一步研究：①对面向商务管理的大数据描述与建模有了一些尝试性的工作，但未能形成一般性的大数据规范化语义描述方法，面向商务管理的大数据系统建模理论仍需深入研究。②面向存储数据的知识发现研究还远没有成熟，传统数据分析方法的基础假设在大数据环境下的适用性问题^[78]，以及面向半结构化、非结构化、稀疏性、不确定性和非对称性等特征的高效数据分析方法有待深入研究。③在流数据的知识发现方面的研究略显不足，基于采样和概要提取的流数据查询和分析方法还存在许多缺陷。

(3) 关于大数据环境下商业情报分析方法方面的研究

商业情报分析(BI&A: business intelligence and analytics)是运用科学的理论、方法和手



段,对大量的(通常是零散、杂乱无章的)商业情报进行收集、加工整理与评价的基础上,透过分析各种信息交织而成的错综复杂的表面现象,把握其本质内容,从而获取对商业活动规律性的认识,为商业活动提供正确决策方案的管理活动^[79]。随着大数据应用的日益普及,商业情报分析也进入了大数据时代,数据库、机器学习、计量经济学、统计、可视化等方法和技术在商业情报分析中被大量使用^[61]。目前,针对大数据环境下的商业情报分析的研究主要聚焦于以下三方面任务,一是针对通用的信息检索方法不能完全满足商业情报检索过程中对时效性和针对性的特殊需求,研究新的商业情报检索方法;二是针对不同商务管理层次对商业情报的需求不同,研究相应的数据挖掘和统计分析方法来发现隐匿于大数据之中的有用知识;三是通过商业情报分析预测未来可能的商业趋势,以促进管理效率的提高,发现新的商业机会。具体研究进展综述如下。

在商业情报检索方法方面,文献[61,80]总结了商务智能和分析(BI&A)领域的技术、工具、数据来源和研究趋势的变化,均指出基于语义本体的商务智能技术是商务分析的关键技术之一。Sheth等^[81,82]提出了一种基于语义本体的信息搜索、整合和分析方法,Azvine等^[83,84]提出了一种基于本体建模的实时商务智能系统,Sell等^[85]建立了一个支持商务智能的语义框架,Spies等^[86]对面向商务管理领域的本体建模方法进行了较为深入的研究,Liyang等^[87]提出了一个商务智能作为服务(BI as a Service)的概念框架。

在商业情报挖掘方法方面,由于数据获取越来越便利,众多学者和企业开始利用网络大规模数据挖掘潜在的商业规律,并在一系列挖掘方法上取得了突破。**对社会媒体观点分析及情感词典构建的研究**,文献[88]首先使用 bootstrapping 方法构建情感词集,然后使用有约束的对称非负矩阵分解算法来自动地给词典中的每个词标注情感极向。文献[89]首先使用标准词典以及语料,通过多种相似性计算方法构建不同的情感词相似图,然后在相似图中执行标签传播算法。**对富媒体情感分析研究**,Baldoni等^[90]以用户备注的标签作为信息源,使用本体驱动做语义 web 上的情感分析,Wu等^[91]针对富媒体数据基于标签和视觉相关性提出了多模态推荐。**对观点挖掘应用于商业领域的研究**,文献[92]通过构建专家系统 WeaknessFinder,采用基于产品特征的情感分析方法帮助制造商从中文评论中挖掘产品的缺陷,从而提高产品质量和竞争力。文献[93]基于潜在狄利克雷分布(LDA),提出了一个抽取消费者满意度的潜在维度以及确定那些维度的情感倾向、标签、有效性、重要性、动态变化以及异构性的统一框架。文献[94]采用一种有约束的优化方法通过在线产品评论挖掘市场结构,利用在线评论中的“优点-缺点”结构来抽取短语,并将这些短语按照语义进行分类。文献[95]将多属性选择模型与多种文本分析方法结合起来研究产品评论对消费者产品选择决定的影响。文献[96]利用词共现关系和语义网络关系从在线消费者评论中挖掘市场结构。文献[97]提出了动态分层线性模型(DHLM)分析在线口碑、广告和品牌绩效的关系。文献[98]通过度量在线优惠券的效果、关键词检索和在线评论研究了营销工作和在线口碑对产品销量的影响。文献[99]构建了一个贝叶斯泛函线性模型,揭示了不同“时刻”评价是如何影响整体评价的。



在商业趋势预测方法方面, Cody 等^[100]提出了一个商业情报和知识管理的集成框架, Sheth 等^[101,102]尝试将语义关联识别和知识发现方法应用于国家安全中和企业风险与合作机会发现中。Bollen 等^[103]以天作为时间轴的基本单位, 使用 POMS 抽取 Twitter 上发布微博的六种情感, 通过统计建模, 分析这些情感波动的出现和股市、石油价格以及社会重大事件之间的关系。使用 Twitter 数据, 来自惠普实验室的 Asur 和 Huberman 成功的通过线性回归预测出了电影的票房, 这比传统的基于市场调查的预测准确性要高^[104]。文献[105]利用大规模网络数据分析股市的行情。文献[106]基于网络大数据(在线社交网络)研究了抗议招募的动态变化规律, 并且分析其对政治选举的影响。文献[107]通过分析 YouTube 上 500 万个视频每日浏览量的时序变化规律得出, 大部分用户行为可以准确地用泊松分布来描述。文献[108]使用电影脚本的浅层文本特征结合自然语言处理技术和统计学习方法来预测在电影投资上的回报。

现状分析: 各种大数据分析 with 计算方法的发展为商业情报分析方法的改进与创新提供了机会, 驱动了商业情报分析方法的发展。通过对大数据环境下的商业情报分析方法的国内外研究现状的总结与分析可以看出, 在商业情报检索方法、社交媒体观点分析及情感词典的构建、观点挖掘应用于商业领域等方面已经取得了一些进展, 但依然还存在着: ①在商业情报检索方法方面, 基于语义本体的信息检索方法是主流, 但缺少针对商业情报分析的语义本体建模方面的研究。②使用大规模流数据实现实时动态的情报检索是商业情报检索的客观需要和研究趋势。③响应实时性需求, 通过对中文主题识别, 基于社交媒体流(流数据形式的多源异构的社交媒体数据)进行动态观点挖掘是研究趋势之一。④在商业趋势预测方面, 只有简单基于统计分析的模型, 缺少有效的基于大数据分析的模型和方法, 在社交媒体观点挖掘的客户倾向性分析方法和商业流行趋势识别方法的基础上, 研究基于多信息源的具有时空域特性的客户倾向性模型和变化趋势预测方法, 研究基于时空分析的商业流行趋势动态演化模型及相应的预测方法是商业情报分析的前沿问题之一。

(4) 关于大数据驱动的个性化用户建模与推荐方法方面的研究

个性化用户建模和推荐方法是目前商务领域的热点研究问题, 大数据环境对个性化的用户建模和推荐产生了重大影响。一方面, 大数据环境下, 支持用户兴趣建模的数据涵盖了消费者在不同时空维度的浏览信息、搜索信息、购买信息、评论信息、社交信息以及消费者情绪信息、地理位置及环境信息等, 有助于提高用户兴趣模型的有效性。另一方面, 大数据的多源性、海量性、异构性、动态性和稀疏性等特点为用户建模和推荐过程中的数据分析、模型构建、推荐算法设计提出了更高的要求。具体研究进展综述如下。

在个性化推荐的多源数据分析方面。早期的个性化推荐系统主要基于在线浏览记录和购买记录等数据构建用户兴趣模型, 在亚马逊等电子商务网站的个性化推荐中得到了广泛应用。随着 Web2.0 等互联网技术的发展, 在线评论、社交网络等新兴信息源在网络用户建模中得到了越来越多的使用。在线评论对用户购买决策和企业收益的影响现有研究已经进行了广泛验证^[109,110]。基于在线评论, 研究者抽取评论中的特征词和情感词, 识别用户所关



注产品的属性及其情感,进而构建用户兴趣模型^[111,112]。在线社交网络是基于兴趣或影响构建的关系网络。大量实证研究表明,与企业推荐相比,朋友推荐的信息或产品更能得到用户的信任^[113]。基于社交网络推荐好友和信息、在个性化推荐中引入社交信息以提高个性化推荐的效果是现有研究的主要思路^[114]。Aral 和 Walker 等人研究了社交网络上个体影响他人和被他人影响的度量方法^[115],该方法为集成社交网络信息进行个性化推荐提供了理论基础。Yang 等人提出了一种基于信任圈的个性化推荐方法,所提方法可以有效利用朋友之间的信任,提高个性化推荐的效果^[116]。上述研究的共同特点是基于某一高价值密度的数据源构建用户兴趣模型。实际上,用户兴趣通常会在不同类型的数据中得到反映,而不同数据源所反映的用户兴趣往往具有一定的相关性。**综合不同数据源的用户兴趣建模是多源数据环境下个性化推荐的重要研究内容。**Ansari 等人综合利用用户陈述偏好、其他消费者和专家的偏好、产品类型和用户特征等信息构建用户兴趣模型^[117]。Pazzani 研究了利用网页内容、用户评分和人口学特征等信息的用户建模方法,对集成多源信息的用户建模问题进行了深入探讨^[118]。He 和 Baltrunas 等人则将个性化推荐的上下文环境集成到用户建模过程,提出了结合上下文感知的个性化推荐方法^[119,120]。

在个性化推荐方法的研究方面。协同过滤、基于内容的过滤是个性化推荐的经典方法。基于余弦相似性、Pearson 相关性和有约束的 Pearson 相关性等相似度计算方法,研究者设计了大量应用于不同领域的协同过滤和基于内容的过滤方法^[121-123]。**随着研究的深入,研究者将贝叶斯理论、图挖掘、马尔可夫模型等方法引入个性化推荐,**对提高个性化推荐的效果起到了积极作用。例如,Huang 等人构建了用户-产品的二部图模型,设计了基于图挖掘的个性化推荐方法^[124]。Moon 和 Russell 等设计了基于联合空间映射和空间选择模型的个性化推荐方法,提高了推荐的准确性^[125]。Ansari 等人设计了基于贝叶斯偏好模型的个性化建模方法以及基于 MCMC 方法的参数估计方法,与协同过滤等传统推荐方法相比,该模型具有更高的推荐准确率^[117]。Sahoo 等构建了基于隐马尔可夫方法的个性化推荐模型,对用户兴趣的时序变化规律进行挖掘^[126]。随着微博等各类社交媒体的广泛应用,**在线社交网络中具有相似兴趣的用户组成不同的兴趣群体,兴趣群体发现、群体兴趣建模以及面向用户群体的个性化推荐是目前推荐领域的热点研究方向。**在兴趣群体发现方面,商集和盖集是目前虚拟社区发现的两类主要思路。商集方法挖掘出的所有社区互不相交^[127,128],而盖集方法获得的社区间的成员可以交叉^[129,130]。在个性化群推荐方面,群推荐列表的产生是群推荐的核心,其关键问题是如何将个体偏好集结为群体偏好,最终产生群推荐结果。目前,集结策略主要包括结果集结、评分集结和模型集结等方法^[131,132]。

在冷启动、稀疏性和多样性等个性化推荐难题的研究方面。由于与新产品和新用户相关的使用数据不足、不同用户的兴趣重叠比例通常较低、以及企业营销的不同目标(增加销量或引导消费),冷启动、稀疏性和多样性是个性化推荐所面对的持续的难题。课题组陈国青教授等人的研究指出,社会化推荐是解决个性化推荐冷启动问题的可靠思路^[133]。**为了缓解稀疏性对用户建模的影响,Sun 等将转移相似度的概念引入个性化用户建模,解决了数**



据稀疏情况下用户相似度计算精度不高的问题^[134]。Huang 等人在基于二部图的协同过滤建
模中引入了传递关联性的概念，有效缓解了数据稀疏性的负面影响^[135]。实证研究表明，**推
荐多样性是影响用户满意度的重要因素^[136-138]**。合理的评判指标是多样性推荐研究的基础。
研究者提出了面向个体用户^[139,140]和用户群体^[141-143]的推荐多样性度量方法。**用户建模阶段
的多样性优化方法和推荐列表生成阶段的多样性优化方法**是目前提升推荐多样性的两种主
要思路。研究者通过保持长尾产品的评分在用户建模阶段保留多样性信息^[143]，并利用主题
多样性优化方法^[136]、二次排序方法^[144]和离散选择模型^[145]等方法在推荐列表生成阶段对候
选推荐产品进行多样性优化。

在推荐结果的可信性和可用性研究方面。从商务管理的角度，生成推荐列表并非推荐
系统的最终目标，说服用户购买或使用推荐的产品、通过推荐系统提高用户满意度、借助
推荐系统提高企业收益具有更为现实的意义。Chen 和 Xie 研究了在线评论对用户购买决策
的影响，为企业利用在线评论引导用户的购买决策提供了理论指导^[146]。课题组的刘业政教
授等结合心理问卷和眼动仪数据，对信息展示页面的可用性问题进行了研究，在个性化的
信息展示方面进行了探索^[147]。课题组的姜元春等研究了考虑用户满意度的个性化推荐方
法，该方法将在线评论引入个性化推荐，提高了推荐产品的用户满意度^[148]。课题组还设计
了基于个性化推荐的动态捆绑策略，通过制定动态的价格折扣吸引用户购买推荐的产品^[149]。
Garfinkel 等考虑消费者的购买费用，设计了以节省最大化为目标的个性化推荐算法^[150,151]。
腾讯公司将社交信息引入个性化推荐，在“应用宝”中产生 APP 推荐列表后，通过显示“***
等好友正在使用该 APP”等信息提高用户使用推荐 APP 的可能性。

现状分析：个性化推荐是目前商务领域的热点研究问题，取得了丰富的研究成果。针
对大数据环境下的个性化用户建模和推荐问题，研究者从用户兴趣模型构建、推荐方法设
计等方面进行了探索，但在以下方面还存在不足：①大数据为个性化推荐的用户建模带来
了许多新的挑战，基于全网数据的用户兴趣图谱构建、基于多源数据的交叉学习以及考虑
时序特征的用户建模等问题的研究还存在不足。②用户群体建模的研究尚处于起步阶段，
用户兴趣社区的划分与识别、用户群体兴趣的集结与演化等问题有待深入研究。③相似度
计算是个性化推荐的永恒问题，大数据环境下相似度计算的稀疏性问题、多目标推荐问题、
以及多源推荐结果的冲突消解问题等有待进一步研究。综合考虑消费者需求与认知、企业
收益与长期目标的个性化推荐方法也需要更为深入的探索。

（5）关于面向商务管理大数据支撑平台的核心理论与关键技术方面的研究

继云计算之后，“大数据”的概念成为了工业界和学术界的热门话题。数据资源的价值逐
步被人们所认知，商务管理的决策依据转向了数据。“让数据说话”已经成为当代最为重要的
IT 命题之一。大数据在规模、增长、分布、处理等方面呈现出的诸多复杂特征，使得传统
技术难以满足新的应用需求，为大数据量身定制一套合适的计算架构成为重要研究难题。
当前国内外针对大数据平台的研究主要集中在对体系结构的改进和优化，涵盖大数据的数
据存储、管理、分析与计算等诸多方面。考虑到人在大数据使用中的重要地位，近几年大



数据平台下人机协作机制也引起相关学者的重视。具体研究进展综述如下。

在大数据支撑平台的体系结构方面。从大数据中获取有价值的知识，对企业或组织的风险评估、品牌管理、运营优化等有重要作用，已经成为众多组织和企业突破外部竞争压力、发展差异化战略的主要途径。然而，现有的商务智能应用往往由不同的独立系统分别完成，每个系统都包含共有的处理步骤，如数据评估、清理、过滤、可视化等。遭遇大规模异构数据时，系统本身的成本以及数据在不同系统间的载入导出成本将显著增加。故构建面向商务管理的大数据支撑平台，设计统一的信息系统体系结构，适应多源异构数据的载入、访问，且灵活地处理不同的商务管理任务显得迫切而必要。文献[152]从大数据处理的不同阶段分析当前大数据处理的主要挑战。为了解决数据在不同存储系统之间移动和载入代价过高的问题，文献[153]构建了基于 HDFS 和 RDBMS 等多存储系统的查询系统 MISO。专利[154]提供了一个统一的方法来管理来自异构数据源的数据，且用户可以通过交互界面进行修改和可视化。在任务集成方面，文献[155]提出了一个六层的私有云平台架构，使用通信队列作为云引擎，通过松弛耦合的方式保持各层之间以相对独立的方式发布和订阅信息，并使用插件式算法框架适应大量的半结构化和非结构化媒体数据。文献[156]论述了第二代 Hadoop 平台的设计、开发和部署，新的架构将解耦程序模型引入底层资源管理、调度功能等。文献[157]使用分层编码来标记 Hadoop 中的模式数据，进而将 RDBMS 和 MapReduce 融合到统一的系统中来进行大数据分析。文献[158]在 Hadoop 平台上研究了面向大数据集的机器学习方法库，可以使用 MapReduce 范式进行大数据应用分析。文献[159]研究构建了分布式机器学习的系统框架，使得用户可以更轻松地将算法推广到更多管理决策问题研究中。

在大数据管理方面。海量性和动态性是大数据应用的基本特征。由于在弹性扩展能力、高并发高通量的存储性能以及超级计算性能等方面的实际需求，大数据存储系统通常构建于云计算平台之上，也就要求其系统设计在保证可靠性的前提下，需要不断优化存储效率和访问效率，提升数据管理的时效性。因此，大数据的有效存储和时效管理是大数据在商务管理中充分应用的重要保证。**对大数据存储方面的研究**，为了适应互联网中呈几何级数爆炸式增长的数据存储和管理的迫切需求，云计算平台上的大数据存储通常采纳以 GFS、HDFS 等为代表的大规模分布式文件系统，这种存储结构具有良好的可扩展能力，但也面临节点经常失效、存储与访问性能一般等问题^[160]。对大数据的存储与布局问题，国内外学者主要研究了基于节点容量、负载能力、负载均衡的数据分配与转移策略，还探索了面向数据密集型应用的数据布局问题。文献[161]针对大规模数据存储中的数据布局难以平衡等问题，运用被动平衡与积极平衡相结合的方法，实现存储节点容量的公平分配；文献[162]为了提高数据密集型应用中数据访问的并发性，研究设计了一种基于管理目标的数据预分配策略；文献[163]分析了数据密集型应用中的传输次数、数据集大小以及带宽等限制性因素，提出了一个三阶段的数据分配策略以优化数据密集型应用的执行效率；文献[164]提出了一种基于数据依赖关系的数据存储文件分组策略；文献[160]则将数据的依赖性定义为可以由用户进行预设置的一个数据属性，进而给出了考虑数据间依赖关系的大规模数据存储方法；



文献[165]讨论了影响存储可靠性的主要因素,并从冗余角度研究了数据存储的可靠性问题。**对数据时效性管理的研究**,具备完整时间戳的实体数据时效性管理相对易于实现,但实际应用中更为常见的时间戳缺失背景下的时效性识别与管理则极为困难。文献[166]通过构造时序模型对实体的时序进行识别;文献[167]通过构造时序图,解决查询相关时效性和用户相关时效性的判定问题;文献[168]研究了基于概率的数据时效性评估问题。

在大数据计算架构方面。当前大数据处理的主流模式一般基于分布式文件系统,将大量数据分散到众多计算节点上,并在云计算环境下采用分布计算和聚合的并行处理方法,充分利用异地分布的各种计算资源,从而全面提升大规模数据的处理速度。目前,大数据的并行处理模型主要有三种:一是在高性能计算上获得广泛应用的 MPI 技术,二是谷歌和雅虎所采用的 Map-Reduce 处理框架,三是微软提出的 Dryad 模型。这些并行处理模型在处理任务的分解策略和任务的部署方式以及任务间的协调机制上存在明显差异。**对处理任务部署的研究**,MPI 模式利用存储节点和计算节点的分离策略提升了数据存储系统的可靠性和管理水平,拓展了并行存储的输入输出带宽,但由于采用移动数据的理念,MPI 需要将数据移动到相应的计算节点,也就相应增加了网络传输的压力。Map-Reduce 和 Dryad 则是基于移动计算的原理,将存储和计算部署在同一节点,避免大量数据的传递,从而减轻网络传输压力。**对处理任务分解的研究**,区别于 MPI 模式落后的人工操作方式,Map-Reduce 和 Dryad 则是将数据划成固定规模的文件块并分散存储于各个集群节点,并将计算任务移动到存储节点或附近完成。对于彼此间不存在数据依赖的单一任务处理,Map-Reduce 和 Dryad 自动地实现任务分解;对于存在依赖关系的复杂任务处理,Map-Reduce 将其转化为多个单一的 MR 任务,运用串联机制完成处理。**对处理任务间协同机制的研究**,MPI 模式是基于特定的消息机制进行任务间的数据交互。而在 Map-Reduce 框架中,Reduce 与 Map 任务之间则通过中间文件进行交流,其协同机制主要是处理任务执行顺序和中间数据的调度优化。类似地,Dryad 模型中任务之间的通信也是通过访问中间数据进行。综上所述不难发现,Map-Reduce 和 Dryad 将计算任务移动到存储节点,更为有利于实现大数据的处理。为了进一步提升数据处理性能,还有学者从调度策略优化^[169]、计算资源分配^[170]、工作流的修改和任务调度^[171-173]等方面研究了 Map-Reduce 的优化技术。以 Map-Reduce 为代表的分布式计算框架已经广泛运用到各个领域,基于特定领域的大数据计算系统也不断提出,定义更加方便的抽象编程和接口说明成为现有的领域导向的大数据系统的重要研究方向^[174]。

在大数据环境下人机协同方面。大数据分析已经为商务活动提供诸多智能的、有价值的信息,很多管理者也意识到数据驱动决策的重要性。例如有研究表明,数据驱动决策者借助大数据分析技术提升了超过 4%的生产力,6%的利润率和 50%的市场价值^[175]。然而在大数据分析过程中,人的主观直觉与大数据分析结果之间如何协调,是一项重要的挑战。近期,Rosenzweig 指出了一系列分析方法的种种弊端并呼吁管理者应当在数据支撑的基础上合理利用自己的主观判断来做决策^[176]。管理者可利用各种方法来做决策,但并没有唯一的最好方法,都需要考虑方法的情境适用性^[177]。例如模型化的分析工具更适合高度结构化



的决策问题，而直觉判断则适用于情境复杂的非结构化问题。大数据分析可以帮助识别新问题亦或是对已有问题寻找解决方案，而人的主观直觉在创造新思想方面尤为重要^[177]。Shvetank Shah 等人研究发现，在商务管理中，如果公司员工不能有效地将数据分析与主观判断相结合，公司对分析工具的投入很可能是无用的甚至是有害的^[178]。他们调研了来自 22 个跨国公司的 5000 多名员工，并将他们分为三组：强烈依赖数据分析的经验主义者、过度依赖主观判断的本能决策者以及可以很好地平衡主观判断和数据分析的怀疑主义者，结果表明，怀疑主义员工在各种绩效指标上都超出其他两组 24%，因而他们呼吁“大数据”应当跟“大判断”紧密结合起来，从而更好地进行商务决策。

现状分析：大数据复杂特征对数据获取与存储、数据组织与管理，数据分析与计算以及数据展示与商务应用都提出了新的挑战。目前，在大数据平台的理论与实践方面已经取得了一定的成效，但其中还存在：①针对大数据环境下体系结构的研究，尚缺乏统一的认识，设计面向商务管理的大数据支撑平台体系结构对整个系统的成功搭建与运行至关重要。②现有的大数据存储策略未能考虑到存储节点的综合性能优化，也忽视了对面向商务管理的冷门数据处理；在大数据治理方面的研究主要集中在技术层面的提升，尚缺乏面向商务管理的大数据时效性机制设计。③Map-Reduce 在大数据处理中的应用已经非常普遍，但其理论上的优化过于集中在并行处理框架的局部算法优化，没有从整体上考虑大数据处理的全局优化。④大数据环境下人机协作的研究越来越得到学者重视，如何处理人的主观判断和数据驱动管理的协同关系将影响着系统平台对商务管理各模块的支撑能力以及整个平台的应用能力。

(6) 关于大数据在商务管理中的应用实践方面的研究

数据是知识经济时代的重要生产要素，是组织获取竞争优势的战略资源之一。在大数据环境下，组织可以利用物联网、互联网等技术手段从生产过程、管理过程、交易过程、社交网络、在线评论等不同途径获取商务管理所需的数据。随着数据规模的快速膨胀、数据来源和结构的异常复杂，如何获取可用数据，从中挖掘有价值的知识，帮助组织发现经营管理的问题并提出相应的解决方案，已经成为大数据环境下组织面临的重要问题^[61]。大数据环境对顾客洞察、市场营销、企业运营、商业模式创新等商务管理过程产生了重要影响，并已经出现了众多的应用。具体研究进展综述如下。

在大数据应用前景的研究方面。大数据在商务管理中的应用，其价值创造主要体现在提高组织与部门间的透明度、发现用户需求与跟踪偏好变化、个性化服务、决策支持以及商业模式、产品和服务创新等五个方面^[179]。例如，**在个性化服务方面**，企业在大数据环境下可以从更多的渠道获得信息，描述用户偏好的信息趋于真实化，企业可以从公开的数据中了解用户行为，从而可推出更精准的个性化产品和服务以最大化企业利益^[180]；**在生产与服务管理方面**，组织在设计自己的产品时，可以通过监测和分析来自社交网络平台上的相关评论信息，发现客户和消费者对产品的实际看法，并将分析结果整合到产品生产决策模型中以精确地预测产品设计对客户流失、消费者忠诚度和满意度的影响^[181]。**在行业应用方**



面,大数据中的大价值将会给消费、旅游、金融、公共服务、医疗卫生、电子商务、交通、制造等众多行业带来机会和挑战^[179,182]。例如,对于消费行业,企业可以通过捕捉互联网上相关的评论、视频、博客、微博等信息来进一步了解消费者,更有效地提供服务、管理品牌、开展营销等;对于金融服务行业,银行、证券、保险等均依赖于对海量数据的集成和分析,实现目标交叉销售,规避欺诈风险;对于交通行业,通过对气象信息、地理信息、车况信息、交通实况信息等进行大数据分析,制定最佳出行方案。

在大数据的实践应用方面。大数据环境下的商务管理创新已经不再只停留在理论研究层面,而是在全球范围内掀起了推进大数据应用的热潮,且已经开始在营销管理、投资决策、产品设计决策、商业模式创新、品牌管理等各种类型的管理实践中发挥重要作用^[183]。

大数据环境下的精准营销和广告投放。eBay 通过从消费者的点击、购买、评论等行为大数据中(100PB/天)挖掘消费者的需求和隐性偏好,为制定更为精准的营销策略提供有效支持。与此同时,eBay 还建设了基于大数据分析的广告投放优化系统,实现对广告投入产出的及时预测,甚至可以精确计算出每一个关键字为 eBay 带来的投资回报,从而在成功降低产品广告投入的同时,维持了较好的企业业绩增长^[184]。

大数据环境下的金融投资。英国金融公司 Derwent Capital 运用大数据技术分析社交平台 Twitter 上的公众信息记录(2.5亿条/天),及时感知并预测社会公众对某一特定股票的看法,从而判断金融市场走向,优化投资决策。此外,考虑到消费者在 Facebook、Twitter、Youtube 等社交网络的行为不仅能够反映其社会关系,更能反映消费者对特定事物的情感^[185],基于社交平台的大数据分析已经被广泛应用于企业的产品推广、客户维持等商务管理实践中^[186]。

大数据环境下的产品创新。Ford 汽车公司正在进行的大数据项目尝试将消费者需求引入汽车产品的开发过程^[187],利用 RFID 等传感装置对汽车生产开发环境的数据进行收集,还对互联网上的博客文章、评论和其他类型数据进行实时跟踪,进而通过这些流数据和非结构化数据的有效利用,以期能够更好地了解汽车运转的状态、消费者使用汽车的方法以及消费者的需求,并将这些信息反馈到汽车产品的生产流程中,优化消费者的体验。

大数据环境下的商业模式创新。2013 年 2 月 21 日苏宁开启“云商”模式,正式实行“店商+电商+零售服务商”的创新商业模式,努力实现实体店面和电子商务的线上线下整合、实体产品和服务产品的产品种类整合、以及苏宁和平台加盟商的企业整合。大数据和云计算为“云商”模式创新提供了极为有力的技术支撑,使其可以综合利用云计算平台上的产品资源、服务资源和数据资源,构建大数据时代在线零售行业的云服务模式,为消费者建立虚实互动的体验模式^[188]。

大数据环境下的品牌管理。很多公司利用大数据平台实时监控社交媒体网站,收集分析数百万的微博、评论等用户在线内容,可以识别消费者对品牌或产品的反应变化并迅速做出市场反应,从而提高品牌价值^[183]。市场营销专家开始重视日益流行的在线社区,开始通过构建、维护和分析社区中消费者的反应,从而加强品牌管理^[189]。例如很多知名品牌入驻在线社区,通过公告牌、论坛、聊天室等社交工具来增强与用户的交流,扩大品牌影响力。

现状分析:大数据环境下商务管理创新的大幕已经在全球范围内拉开,数以万计的组



织和企业已经开始部署大数据应用以创新发展模式、提升核心竞争力^[190]。然而,面向商务管理的大数据研究还远没有成熟,利用大数据推动商务管理创新还遭遇到数据获取、存储、计算、分析、集成、可视化等方法技术问题。从大数据在商务管理方面的应用实践回顾不难发现,虽然大数据在企业经营实践中已经有了成功的应用,但现有的实践多是某一个或一类大数据技术的尝试性应用,有必要结合我国企业的管理需求,系统地运用大数据环境下商务管理中顾客洞察与市场营销策略、企业运营与优化、商务模式创新以及商务分析与计算方法等方面的研究成果,全面提升我国企业的商务管理水平。

1.3 参考文献

- [1] D. Barton, Making advanced analytics work for you, *Harv. Bus. Rev.*, 90 (2012) 78-83.
- [2] 李国杰, 大数据研究的科学价值, *中国计算机学会通讯* 8(9) (2012) 8-15.
- [3] N. Sawant, H. Shah, Big Data Ingestion and Streaming Patterns, in *Big Data Application Architecture Q & A* Springer, 2013.
- [4] National Research Council, *Frontiers in Massive Data Analysis*. Washington, DC: National Academies Press, 2013.
- [5] A. TITIRISCA, ETL as a Necessity for Business Architectures, *Database Systems Journal* 4(2) (2013) 3-12.
- [6] A. Rajpurohit. Big data for business managers—Bridging the gap between potential and value, in *Big Data*, 2013 IEEE International Conference on, 2013.
- [7] I. H. Witten, E. Frank, A. Mark, *Data Mining: Practical machine learning tools and techniques* (3rd Edition): Morgan Kaufmann, San Francisco. Retrieved, 2011.
- [8] B. Buelens, P. Daas, J. Burger, M. Puts, J. van den Brakel, Selectivity of Big data, *Statistics Netherlands* (2014).
- [9] N. G. Weiskopf, C. Weng, Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research, *Journal of the American Medical Informatics Association* 20(1) (2013) 144-151.
- [10] L. L. Pipino, Y. W. Lee, R. Y. Wang, Data quality assessment, *Communications of the ACM* 45(4) (2002) 211-218.
- [11] R. Y. Wang, H. B. Kon, S. E. Madnick. Data quality requirements analysis and modeling, in *Data Engineering*, 1993. Proceedings. Ninth International Conference on, 1993.
- [12] J. Michnik, M.-C. Lo, The assessment of the information quality with the aid of multiple criteria analysis, *European Journal of Operational Research* 195(3) (2009) 850-856.
- [13] C. Moraga, M. Moraga, C. Calero, A. Caro. SQuaRE-aligned data quality model for web portals, in *Quality Software*, 2009. QSIC'09. 9th International Conference on, 2009.
- [14] E. M. Pierce, Assessing data quality with control matrices, *Communications of the ACM* 47(2) (2004) 82-86.
- [15] C. Batini, F. Cabitza, C. Cappiello, C. Francalanci, A comprehensive data quality methodology for web and structured data, *International Journal of Innovative Computing and Applications* 1(3) (2008) 205-218.
- [16] C. Bizer, P. Boncz, M. L. Brodie, O. Erling, The meaningful use of big data: four perspectives--four challenges, *ACM SIGMOD Record* 40(4) (2012) 56-60.
- [17] J. Bleiholder, F. Naumann, Data fusion, *ACM Computing Surveys (CSUR)* 41(1) (2008) 1.



- [18] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, F. Silva, Directed diffusion for wireless sensor networking, *Networking, IEEE/ACM Transactions on* 11(1) (2003) 2-16.
- [19] O. Younis, S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks, *Mobile Computing, IEEE Transactions on* 3(4) (2004) 366-379.
- [20] S. Roy, M. Conti, S. Setia, S. Jajodia, Secure data aggregation in wireless sensor networks, *Information Forensics and Security, IEEE Transactions on* 7(3) (2012) 1040-1052.
- [21] B. Krishnamachari, D. Estrin, S. Wicker. The impact of data aggregation in wireless sensor networks, in *Distributed Computing Systems Workshops, 2002. Proceedings. 22nd International Conference on*, 2002.
- [22] M. C. Vuran, I. F. Akyildiz, Spatial correlation-based collaborative medium access control in wireless sensor networks, *Networking, IEEE/ACM Transactions on* 14(2) (2006) 316-329.
- [23] J. Zhao, Z. Wei, Z. Li, X. Sun, An adaptive MAC protocol based on spatial and temporal correlation in wireless sensor networks, *Journal of Communications* 6(3) (2011) 262-269.
- [24] G. Zheng, S. Tang, Spatial correlation-based MAC protocol for event-driven wireless sensor networks, *Journal of Networks* 6(1) (2011) 121-128.
- [25] M. Melenhorst, M. Grootveld, M. van Setten, M. Veenstra. Tag-based information retrieval of video content, in *Proceedings of the 1st international conference on Designing interactive user experiences for TV and video*, 2008.
- [26] V. S. Tseng, J.-H. Su, B.-W. Wang, Y.-M. Lin. Web image annotation by fusing visual features and textual information, in *Proceedings of the 2007 ACM symposium on Applied computing*, 2007.
- [27] V. S. Tseng, J.-H. Su, J.-H. Huang, C.-J. Chen, Integrated mining of visual features, speech features, and frequent patterns for semantic video annotation, *Multimedia, IEEE Transactions on* 10(2) (2008) 260-267.
- [28] A. Inselberg, The plane with parallel coordinates, *The Visual Computer* 1(2) (1985) 69-91.
- [29] E. J. Wegman, Hyperdimensional data analysis using parallel coordinates, *Journal of the American Statistical Association* 85(411) (1990) 664-675.
- [30] A. Inselberg, B. Dimsdale, Parallel coordinates, in *Human-Machine Interactive Systems* Springer, 1991.
- [31] Y.-H. Fua, M. O. Ward, E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets, in *Proceedings of the conference on Visualization'99: celebrating ten years*, 1999.
- [32] H. Zhou, X. Yuan, H. Qu, W. Cui, B. Chen. Visual clustering in parallel coordinates, in *Computer Graphics Forum*, 2008.
- [33] N. Andrienko, G. Andrienko, P. Gatalsky, Exploratory spatio-temporal visualization: an analytical review, *Journal of Visual Languages & Computing* 14(6) (2003) 503-541.
- [34] J. Choo, H. Park, Customizing computational methods for visual analytics with big data, *Computer Graphics and Applications, IEEE* 33(4) (2013) 22-28.
- [35] K. Reda, A. Febretti, A. Knoll, J. Aurisano, J. Leigh, A. E. Johnson, M. E. Papka, M. Hereld, Visualizing large, heterogeneous data in hybrid-reality environments, *IEEE computer graphics and applications* 33(4) (2013) 38-48.
- [36] S. Ribaric, T. Hrkac, A model of fuzzy spatio-temporal knowledge representation and reasoning based on high-level Petri nets, *Information systems* 37(3) (2012) 238-256.
- [37] A. Slingsby, J. Dykes, J. Wood, Using treemaps for variable selection in spatio-temporal visualisation, *Information Visualization* 7(3-4) (2008) 210-224.
- [38] D. P. Tegarden, Business information visualization, *Communications of the AIS* 1(1es) (1999) 4.



- [39] 吴佳鑫, Web 环境下信息可视化模型研究[D], 2009, 武汉大学.
- [40] 李昕, 四面体数据高效可视化技术研究[D], 2013, 浙江大学.
- [41] D. Bollegala, Y. Matsuo, M. Ishizuka, A web search engine-based approach to measure semantic similarity between words, *Knowledge and Data Engineering, IEEE Transactions on* 23(7) (2011) 977-990.
- [42] M. Krötzsch, F. Maier, A. Krisnadhi, P. Hitzler. A better uncle for OWL: Nominal schemas for integrating rules and ontologies, in *Proceedings of the 20th international conference on World wide web*, 2011.
- [43] L. Sweeney, k-anonymity: A model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(05) (2002) 557-570.
- [44] C. Garcia-Alvarado, C. Ordonez. ONTOCUBO: cube-based ontology construction and exploration, in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014.
- [45] S. Mukherjee, J. Ajmera, S. Joshi. Unsupervised approach for shallow domain ontology construction from corpus, in *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, 2014.
- [46] A. Maedche, S. Staab, Ontology learning for the semantic web, *IEEE Intelligent systems* 16(2) (2001) 72-79.
- [47] L. V. Lakshmanan, J. Pei, Y. Zhao. QC-Trees: An efficient summary structure for semantic OLAP, in *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, 2003.
- [48] J. Feldman, S. Muthukrishnan, A. Sidiropoulos, C. Stein, Z. Svitkina, On distributing symmetric streaming computations, *ACM Transactions on Algorithms (TALG)* 6(4) (2010) 66.
- [49] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski, C. Kozyrakis. Evaluating mapreduce for multi-core and multiprocessor systems, in *High Performance Computer Architecture, 2007. HPCA 2007. IEEE 13th International Symposium on*, 2007.
- [50] K. Murthy, P. M. Deshpande, A. Dey, R. Halasipuram, M. Mohania, P. Deepak, J. Reed, S. Schumacher, Exploiting evidence from unstructured data to enhance master data management, *Proceedings of the VLDB Endowment* 5(12) (2012) 1862-1873.
- [51] G. Hinton, R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, *Science* 313(5786) (2006) 504-507.
- [52] S. Shalev-Shwartz, Y. Singer, N. Srebro, A. Cotter, Pegasos: Primal estimated sub-gradient solver for svm, *Mathematical programming* 127(1) (2011) 3-30.
- [53] A. R. Neto, G. A. Barreto, Opposite maps: Vector quantization algorithms for building reduced-set svm and lssvm classifiers, *Neural processing letters* 37(1) (2013) 3-19.
- [54] 孟祥福, 马宗民, 张霄雁, 王星, 基于改进决策树算法的 Web 数据库查询结果自动分类方法, *计算机研究与发展* 49(12) (2013) 2656-2670.
- [55] U. Bhowan, M. Johnston, M. Zhang, X. Yao, Evolving diverse ensembles using genetic programming for classification with unbalanced data, *Evolutionary Computation, IEEE Transactions on* 17(3) (2013) 368-386.
- [56] W.-Y. Chen, Y. Song, H. Bai, C.-J. Lin, E. Y. Chang, Parallel spectral clustering in distributed systems, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33(3) (2011) 568-586.
- [57] Z. A. Zhu, S. Lattanzi, V. Mirrokni. A local algorithm for finding well-connected clusters, in *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 2013.
- [58] D. A. Spielman, S.-H. Teng, A local clustering algorithm for massive graphs and its application to



- nearly linear time graph partitioning, *SIAM Journal on Computing* 42(1) (2013) 1-26.
- [59] X. Peng, L. Zhang, Z. Yi. Scalable sparse subspace clustering, in *Computer Vision and Pattern Recognition (CVPR)*, 2013 IEEE Conference on, 2013.
- [60] A. Rodriguez, A. Laio, Clustering by fast search and find of density peaks, *Science* 344(6191) (2014) 1492-1496.
- [61] H. Chen, R. H. Chiang, V. C. Storey, *Business Intelligence and Analytics: From Big Data to Big Impact*, *Mis Quarterly* 36(4) (2012) 1165-1188.
- [62] Y. Zhao, E. Levina, J. Zhu, Community extraction for social networks, *Proceedings of the National Academy of Sciences* 108(18) (2011) 7321-7326.
- [63] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, P. Zhang, Spectral redemption in clustering sparse networks, *Proceedings of the National Academy of Sciences* 110(52) (2013) 20935-20940.
- [64] P. K. Gopalan, D. M. Blei, Efficient discovery of overlapping communities in massive networks, *Proceedings of the National Academy of Sciences* 110(36) (2013) 14534-14539.
- [65] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, P. C. Sabeti, Detecting novel associations in large data sets, *science* 334(6062) (2011) 1518-1524.
- [66] E. Baralis, L. Cagliero, T. Cerquitelli, P. Garza, Generalized association rule mining with constraints, *Information Sciences* 194 (2012) 68-84.
- [67] V. Nebot, R. Berlanga, Finding association rules in semantic web data, *Knowledge-Based Systems* 25(1) (2012) 51-62.
- [68] K.-T. Chuang, H.-L. Chen, M.-S. Chen, Feature-preserved sampling over streaming data, *ACM Transactions on Knowledge Discovery from Data (TKDD)* 2(4) (2009) 15.
- [69] 秦林新, 刘奇志, 一种乱序数据流上的偏倚抽样算法, *计算机研究与发展* 48(z2) (2011).
- [70] G. Cormode, S. Muthukrishnan, K. Yi, Q. Zhang, Continuous sampling from distributed streams, *Journal of the ACM (JACM)* 59(2) (2012) 10.
- [71] B. Babcock, M. Datar, R. Motwani. Sampling from a moving window over streaming data, in *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, 2002.
- [72] H. Wang, W. Fan, P. S. Yu, J. Han. Mining concept-drifting data streams using ensemble classifiers, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003.
- [73] S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, Y.-K. Lee, Sliding window-based frequent pattern mining over data streams, *Information sciences* 179(22) (2009) 3843-3865.
- [74] 李国徽, 陈辉, 挖掘数据流任意滑动时间窗口内频繁模式, *软件学报* 19(10) (2008) 2585-2596.
- [75] F. Cao, J. Liang, L. Bai, X. Zhao, C. Dang, A framework for clustering categorical time-evolving data, *Fuzzy Systems, IEEE Transactions on* 18(5) (2010) 872-882.
- [76] G. Ditzler, R. Polikar, Incremental learning of concept drift from streaming imbalanced data, *Knowledge and Data Engineering, IEEE Transactions on* 25(10) (2013) 2283-2301.
- [77] R. Talmon, R. R. Coifman, Empirical intrinsic geometry for nonlinear modeling and time series filtering, *Proceedings of the National Academy of Sciences* 110(31) (2013) 12535-12540.
- [78] D. Spiegelhalter, The future lies in uncertainty, *Science* 345(6194) (2014) 264-265.
- [79] Wikipedia. <http://zh.wikipedia.org/wiki/情报分析>. 2014-06-28.
- [80] R. T. Ng, P. C. Arocena, D. Barbosa, G. Carenini, J. Gomes, Luiz, S. Jou, R. A. Leung, E. Milios,



- R. J. Miller, J. Mylopoulos, Perspectives on Business Intelligence, *Synthesis Lectures on Data Management* 5(1) (2013) 1-163.
- [81] A. P. Sheth, C. Ramakrishnan, Semantic (Web) Technology In Action: Ontology Driven Information Systems for Search, Integration and Analysis, *IEEE Data Eng. Bull.* 26(4) (2003) 40-48.
- [82] A. Sheth. From semantic search & integration to analytics, in *Semantic Interoperability and Integration*, 2005.
- [83] B. Azvine, Z. Cui, D. Nauck, Towards real-time business intelligence, *BT Technology Journal* 23(3) (2005) 214-225.
- [84] B. Azvine, Z. Cui, D. Nauck, B. Majeed. Real time business intelligence for the adaptive enterprise, in *The 8th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, 2006.
- [85] D. Sell, D. C. da Silva, F. D. Beppler, M. Napoli, F. B. Ghisi, R. Pacheco, J. L. Todesco. SBI: a semantic framework to support business intelligence, in *Proceedings of the first international workshop on Ontology-supported business intelligence*, 2008.
- [86] M. Spies, An ontology modelling perspective on business reporting, *Information Systems* 35(4) (2010) 404-416.
- [87] T. Liyang, N. Zhiwei, W. Zhangjun, W. Li. A conceptual framework for business intelligence as a service (saas bi), in *Intelligent Computation Technology and Automation (ICICTA)*, 2011 International Conference on, 2011.
- [88] W. Peng, D. H. Park, Generate adjective sentiment dictionary for social media sentiment analysis using constrained nonnegative matrix factorization, *Urbana* 51 (2011) 61801.
- [89] G. Xu, X. Meng, H. Wang. Build Chinese emotion lexicons using a graph-based algorithm and multiple resources, in *Proceedings of the 23rd International Conference on Computational Linguistics*, 2010.
- [90] M. Baldoni, C. Baroglio, V. Patti, P. Rena, From tags to emotions: Ontology-driven sentiment analysis in the social semantic web, *Intelligenza Artificiale* 6(1) (2012) 41-54.
- [91] L. Wu, L. Yang, N. Yu, X.-S. Hua. Learning to tag, in *Proceedings of the 18th international conference on World wide web*, 2009.
- [92] W. Zhang, H. Xu, W. Wan, Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis, *Expert Systems with Applications* 39(11) (2012) 10283-10291.
- [93] S. Tirunillai, G. Tellis, Mining Marketing Meaning from Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation, *Journal of Marketing Research* (2014).
- [94] T. Y. Lee, E. T. Bradlow, Automated marketing research using online customer reviews, *Journal of Marketing Research* 48(5) (2011) 881-894.
- [95] N. Archak, A. Ghose, P. G. Ipeirotis, Deriving the pricing power of product features by mining consumer reviews, *Management Science*, 57(8) (2011) 1485-1509.
- [96] O. Netzer, R. Feldman, J. Goldenberg, M. Fresko, Mine your own business: Market-structure surveillance through text mining, *Marketing Science* 31(3) (2012) 521-543.
- [97] S. Gopinath, J. S. Thomas, L. Krishnamurthi, Investigating the Relationship Between the Content of Online Word of Mouth, Advertising, and Brand Performance, *Marketing Science* 33(2) (2014) 241-258.
- [98] X. Lu, S. Ba, L. Huang, Y. Feng, Promotional Marketing or Word-of-Mouth? Evidence from



- Online Restaurant Reviews, *Information Systems Research* 24(3) (2013) 596-612.
- [99] S. K. Hui, T. Meyvis, H. Assael, Analyzing Moment-to-Moment Data Using a Bayesian Functional Linear Model: Application to TV Show Pilot Testing, *Marketing Science* 33(2) (2014) 222-240.
- [100] W. F. Cody, J. T. Kreulen, V. Krishna, W. S. Spangler, The integration of business intelligence and knowledge management, *IBM systems journal* 41(4) (2002) 697-713.
- [101] A. Sheth, B. Aleman-Meza, I. B. Arpinar, C. Bertram, Y. Warke, C. Ramakrishnan, C. Halaschek, K. Anyanwu, D. Avant, F. S. Arpinar, Semantic association identification and knowledge discovery for national security applications, *Journal of Database Management (JDM)* 16(1) (2005) 33-53.
- [102] A. Sheth, Enterprise applications of semantic Web: The sweet spot of risk and compliance, in *Industrial Applications of Semantic Web* Springer, 2005.
- [103] J. Bollen, H. Mao, X. Zeng, Twitter mood predicts the stock market, *Journal of Computational Science* 2(1) (2011) 1-8.
- [104] S. Asur, B. A. Huberman. Predicting the future with social media, in *Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2010 IEEE/WIC/ACM International Conference on, 2010.
- [105] R. M. May, S. A. Levin, G. Sugihara, Complex systems: Ecology for bankers, *Nature* 451(7181) (2008) 893-895.
- [106] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, Y. Moreno, The dynamics of protest recruitment through an online network, *Scientific reports* 1 (2011).
- [107] R. Crane, D. Sornette, Robust dynamic classes revealed by measuring the response function of a social system, *Proceedings of the National Academy of Sciences* 105(41) (2008) 15649-15653.
- [108] J. Eliashberg, S. K. Hui, Z. J. Zhang, From story line to box office: A new approach for green-lighting movie scripts, *Management Science*, 53(6) (2007) 881-893.
- [109] N. Hu, N. S. Koh, S. K. Reddy, Ratings lead you to the product, reviews help you clinch it? The mediating role of online review sentiments on product sales, *Decision Support Systems* 57 (2014) 42-53.
- [110] S. M. Mudambi, D. Schuff, What makes a helpful online review? A study of customer reviews on Amazon. com, *Management Information Systems Quarterly* 34(1) (2010) 11.
- [111] H. Liu, J. He, T. Wang, W. Song, X. Du, Combining user preferences and user opinions for accurate recommendation, *Electronic Commerce Research and Applications* 12(1) (2013) 14-23.
- [112] R. Decker, M. Trusov, Estimating aggregate consumer preferences from online product reviews, *International Journal of Research in Marketing* 27(4) (2010) 293-307.
- [113] H. Ma, D. Zhou, C. Liu, M. R. Lyu, I. King. Recommender systems with social regularization, in *Proceedings of the fourth ACM international conference on Web search and data mining*, 2011.
- [114] X. Yang, Y. Guo, Y. Liu, Bayesian-Inference-Based Recommendation in Online Social Networks, *Parallel and Distributed Systems, IEEE Transactions on* 24(4) (2013) 642-651.
- [115] S. Aral, D. Walker, Identifying influential and susceptible members of social networks, *Science* 337(6092) (2012) 337-341.
- [116] X. Yang, H. Steck, Y. Liu. Circle-based recommendation in online social networks, in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012.
- [117] A. Ansari, S. Essegai, R. Kohli, Internet recommendation systems, *Journal of Marketing Research* 37(3) (2000) 363-375.



- [118] M. J. Pazzani, A framework for collaborative, content-based and demographic filtering, *Artificial Intelligence Review* 13(5-6) (1999) 393-408.
- [119] Q. He, J. Pei, D. Kifer, P. Mitra, L. Giles. Context-aware citation recommendation, in *Proceedings of the 19th international conference on World wide web*, 2010.
- [120] L. Baltrunas, B. Ludwig, F. Ricci. Matrix factorization techniques for context aware recommendation, in *Proceedings of the fifth ACM conference on Recommender systems*, 2011.
- [121] P. Resnick, H. R. Varian, Recommender systems, *Communications of the ACM* 40(3) (1997) 56-58.
- [122] B. Sarwar, G. Karypis, J. Konstan, J. Riedl. Item-based collaborative filtering recommendation algorithms, in *Proceedings of the 10th international conference on World Wide Web*, 2001.
- [123] M. Balabanović, Y. Shoham, Fab: content-based, collaborative recommendation, *Communications of the ACM* 40(3) (1997) 66-72.
- [124] Z. Huang, D. D. Zeng, H. Chen, Analyzing consumer-product graphs: Empirical findings and applications in recommender systems, *Management science* 53(7) (2007) 1146-1164.
- [125] S. Moon, G. J. Russell, Predicting product purchase from inferred customer similarity: An autologistic model approach, *Management Science* 54(1) (2008) 71-82.
- [126] N. Sahoo, P. V. Singh, T. Mukhopadhyay, A hidden Markov model for collaborative filtering, *MIS Quarterly* 36(4) (2012) 1329-1356.
- [127] M. E. Newman, M. Girvan, Finding and evaluating community structure in networks, *Physical review E* 69(2) (2004) 026113.
- [128] M. E. Newman, Modularity and community structure in networks, *Proceedings of the National Academy of Sciences* 103(23) (2006) 8577-8582.
- [129] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, *Nature* 435(7043) (2005) 814-818.
- [130] Y.-Y. Ahn, J. P. Bagrow, S. Lehmann, Link communities reveal multiscale complexity in networks, *Nature* 466(7307) (2010) 761-764.
- [131] J. P. Baskin, S. Krishnamurthi. Preference aggregation in group recommender systems for committee decision-making, in *Proceedings of the third ACM conference on Recommender systems*, 2009.
- [132] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, C. Yu, Group recommendation: Semantics and efficiency, *Proceedings of the VLDB Endowment* 2(1) (2009) 754-765.
- [133] E.-P. Lim, H. Chen, G. Chen, Business intelligence and analytics: research directions, *ACM Transactions on Management Information Systems (TMIS)* 3(4) (2013) 17.
- [134] D. Sun, T. Zhou, J.-G. Liu, R.-R. Liu, C.-X. Jia, B.-H. Wang, Information filtering based on transferring similarity, *Physical Review E* 80(1) (2009) 017101.
- [135] Z. Huang, H. Chen, D. Zeng, Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering, *ACM Transactions on Information Systems (TOIS)* 22(1) (2004) 116-142.
- [136] C.-N. Ziegler, S. M. McNee, J. A. Konstan, G. Lausen. Improving recommendation lists through topic diversification, in *Proceedings of the 14th international conference on World Wide Web*, 2005.
- [137] E. Brynjolfsson, Y. Hu, D. Simester, Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales, *Management Science* 57(8) (2011) 1373-1386.
- [138] E. Brynjolfsson, Y. Hu, M. D. Smith, Consumer surplus in the digital economy: Estimating the



- value of increased product variety at online booksellers, *Management Science* 49(11) (2003) 1580-1596.
- [139] M. Zhang, N. Hurley. Avoiding monotony: improving the diversity of recommendation lists, in *Proceedings of the 2008 ACM conference on Recommender systems*, 2008.
- [140] R. Hu, P. Pu. Enhancing recommendation diversity with organization interfaces, in *Proceedings of the 16th international conference on Intelligent user interfaces*, 2011.
- [141] G. Adomavicius, Y. Kwon, Improving aggregate recommendation diversity using ranking-based techniques, *Knowledge and Data Engineering, IEEE Transactions on* 24(5) (2012) 896-911.
- [142] H. Kim, J. Kim, Y. Ryu. A local scoring model for recommendation, in *Proc. of the 20th Workshop on Information Technologies and Systems (WITS'10)*, 2010.
- [143] M. Levy, K. Bosteels. Music recommendation and the long tail, in *1st Workshop On Music Recommendation And Discovery (WOMRAD)*, ACM RecSys, 2010, Barcelona, Spain, 2010.
- [144] G. Adomavicius, Y. Kwon, Optimization-Based Approaches for Maximizing Aggregate Recommendation Diversity, *INFORMS Journal on Computing* 26(2) (2014) 351-369.
- [145] H. Jiang, X. Qi, H. Sun, Choice-Based Recommender Systems: A Unified Approach to Achieving Relevancy and Diversity, *Operations Research* (2014) 140702093942008.
- [146] Y. Chen, J. Xie, Online consumer review: Word-of-mouth as a new element of marketing communication mix, *Management Science* 54(3) (2008) 477-491.
- [147] 刘业政, 叶俊杰, 蒋玮, 徐俊芬, 基于图上随机游动的 Web 页面布局可用性评价, *管理科学学报* 16(9) (2013) 75-81.
- [148] Y. Jiang, J. Shang, Y. Liu, Maximizing customer satisfaction through an online recommendation system: A novel associative classification model, *Decision Support Systems* 48(3) (2010) 470-479.
- [149] Y. Jiang, J. Shang, C. F. Kemerer, Y. Liu, Optimizing e-tailer profits and customer savings: Pricing multistage customized online bundles, *Marketing Science* 30(4) (2011) 737-752.
- [150] R. Garfinkel, R. Gopal, A. Tripathi, F. Yin, Design of a shopbot and recommender system for bundle purchases, *Decision Support Systems* 42(3) (2006) 1974-1986.
- [151] R. Garfinkel, R. Gopal, B. Pathak, F. Yin, Shopbot 2.0: Integrating recommendations and promotions with comparison shopping, *Decision Support Systems* 46(1) (2008) 61-69.
- [152] A. Labrinidis, H. Jagadish, Challenges and opportunities with big data, *Proceedings of the VLDB Endowment* 5(12) (2012) 2032-2033.
- [153] J. LeFevre, J. Sankaranarayanan, H. Hacigumus, J. Tatemura, N. Polyzotis, M. J. Carey. MISO: Souping up big data query processing with a multistore system, in *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, 2014.
- [154] V. Chandrasekhara, A. Eberhart, S. Kraus, U. Walther, Collaborative workbench for managing data from heterogeneous sources. 2013, Google Patents.
- [155] C. He, X. Fan, Y. Li, Toward ubiquitous healthcare services with a novel efficient cloud platform, *Biomedical Engineering, IEEE Transactions on* 60(1) (2013) 230-234.
- [156] V. K. Vavilapalli, A. C. Murthy, C. Douglas, S. Agarwal, M. Konar, R. Evans, T. Graves, J. Lowe, H. Shah, S. Seth. Apache hadoop yarn: Yet another resource negotiator, in *Proceedings of the 4th annual Symposium on Cloud Computing*, 2013.
- [157] X. Qin, H. Wang, F. Li, B. Zhou, Y. Cao, C. Li, H. Chen, X. Zhou, X. Du, S. Wang. Beyond Simple Integration of RDBMS and MapReduce--Paving the Way toward a Unified System for Big Data Analytics: Vision and Progress, in *Cloud and Green Computing (CGC)*, 2012 Second International Conference on, 2012.



- [158] D. Wegener, M. Mock, D. Adranale, S. Wrobel. Toolkit-based high-performance Data Mining of large Data on MapReduce Clusters, in Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on, 2009.
- [159] T. Kraska, A. Talwalkar, J. C. Duchi, R. Griffith, M. J. Franklin, M. I. Jordan. MLbase: A Distributed Machine-learning System, in CIDR, 2013.
- [160] G. Fedak, H. He, F. Cappello. BitDew: a programmable environment for large-scale data management and distribution, in High Performance Computing, Networking, Storage and Analysis, 2008. SC 2008. International Conference for, 2008.
- [161] B. Welch, M. Unangst, Z. Abbasi, G. A. Gibson, B. Mueller, J. Small, J. Zelenka, B. Zhou. Scalable Performance of the Panasas Parallel File System, in FAST, 2008.
- [162] D. Yoo, K. M. Sim, A Locality Enhanced Scheduling Method for Multiple MapReduce Jobs In a Workflow Application.
- [163] 郑湃, 崔立真, 王海洋, 徐猛, 云计算环境下面向数据密集型应用的数据布局策略与方法, 计算机学报 33(8) (2010) 1472-1480.
- [164] S. Doraimani, A. Iamnitchi. File grouping for scientific data management: lessons from experimenting with real traces, in Proceedings of the 17th international symposium on High performance distributed computing, 2008.
- [165] 穆飞, 薛巍, 舒继武, 郑纬民, 一种面向大规模副本存储系统的可靠性模型, 计算机研究与发展 (5) (2009) 756-761.
- [166] W. Fan, F. Geerts, J. Wijsen, Determining the currency of data, ACM Transactions on Database Systems (TODS) 37(4) (2012) 25.
- [167] 李默涵, 李建中, 高宏, 数据时效性判定问题的求解算法, 计算机学报 35(11) (2012) 2348-2360.
- [168] B. Heinrich, M. Klier, A novel data quality metric for timeliness considering supplemental data, (2009).
- [169] S. Tayal, Tasks scheduling optimization for the cloud computing systems, International Journal of Advanced Engineering Sciences And Technologies (IJAESt) 5(2) (2011) 111-115.
- [170] T. Sandholm, K. Lai, MapReduce optimization using regulated dynamic prioritization, ACM SIGMETRICS Performance Evaluation Review 37(1) (2009) 299-310.
- [171] K. Kambatla, N. Rapolu, S. Jagannathan, A. Grama. Asynchronous algorithms in mapreduce, in Cluster Computing (CLUSTER), 2010 IEEE International Conference on, 2010.
- [172] W. F. Boyer, G. S. Hura, Non-evolutionary algorithm for scheduling dependent tasks in distributed heterogeneous computing environments, Journal of Parallel and Distributed Computing 65(9) (2005) 1035-1046.
- [173] J. Polo, D. Carrera, Y. Becerra, V. Beltran, J. Torres, E. Ayguadé Performance management of accelerated mapreduce workloads in heterogeneous clusters, in Parallel Processing (ICPP), 2010 39th International Conference on, 2010.
- [174] S. Sakr, A. Liu, A. G. Fayoumi, The family of MapReduce and large-scale data processing systems, ACM Computing Surveys (CSUR) 46(1) (2013) 11.
- [175] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, D. Barton, Big Data, The management revolution. Harvard Bus Rev 90(10) (2012) 61-67.
- [176] P. Rosenzweig, Left Brain, Right Stuff: How Leaders Make Winning Decisions: Profile Books, 2014.
- [177] P. M. Tingling, M. J. Brydon, Is decision-based evidence making necessarily bad, MIT Sloan



- Manage. Rev. 51(4) (2010) 71-76.
- [178] S. Shah, A. Horne, J. Capellá, Good data won't guarantee good decisions, Harv. Bus. Rev. 90(4) (2012) 23-25.
- [179] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers, Big data: The next frontier for innovation, competition, and productivity, (2011).
- [180] 孟小峰, 慈祥, 大数据管理: 概念, 技术与挑战, 计算机研究与发展 50(1) (2013) 146-169.
- [181] S. Parise, B. Iyer, D. Vesset, Four Strategies to Capture and Create Value from Big Data, (2012).
- [182] A. J. Severin, Dealing with data: training new scientists, Science (New York, NY) 331(6024) (2011) 1516-1516.
- [183] G. George, M. R. Haas, A. Pentland, Big Data and Management, Academy of management Journal 57(2) (2014) 321-326.
- [184] Uwe Mayer, Nish Parikh, Gyanit Singh, Large-Scale Clickstream and transaction log mining in practice, ebay Report, October, 2013.
- [185] J. Brown, A. J. Broderick, N. Lee, Word of mouth communication within online communities: Conceptualizing the online social network, Journal of interactive marketing 21(3) (2007) 2-20.
- [186] A. S. Acar, M. Polonsky, Online social networks and insights into marketing communications, Journal of Internet Commerce 6(4) (2007) 55-72.
- [187] J. Hiner. Ford's Big Data chief sees massive possibilities, but the tools need work. 2012 <http://www.zdnet.com/fords-big-data-chief-sees-massive-possibilities-but-the-tools-need-work-700000322/>.
- [188] 新浪科技. 苏宁详解云商集团组织架构调整. <http://tech.sina.com.cn/i/2013-02-21/14138077948.shtml>
- [189] G. McWilliam, Building stronger brands through online communities, Sloan Management (2012).
- [190] J. Bughin, M. Chui, J. Manyika, Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, McKinsey Quarterly 56(1) (2010) 75-86.

2、课题研究内容、拟解决的关键科学问题、具体研究目标及在重大项目中所起的作用（此部分为重点阐述内容）。

2.1 研究目标

根据大数据环境下商务管理的发展需求和《重大项目指南》的要求, 确定本课题(课题五)研究的总体目标为: 系统研究面向商务管理的大数据分析和计算的核心方法以及关键技术, 在面向商务管理的大数据可用性评估方法、面向商务管理的超高维稀疏性大数据快速分析方法、面向商务管理的中文主题识别方法、个性化信息推荐的多目标优化方法、大数据支撑平台体系结构设计理论等关键科学和技术问题上取得突破, 构建面向商务管理的内外部数据获取系统、计算实验平台以及原型系统, 争取获得能够指导我国企业实现大数据环境下商务管理创新实践、具有一定国际影响力的系统性学术成果, 促进相关科学领域的发展, 提升我国商务管理理论的研究水平, 并建设一支具有国际影响力和竞争力的研究队伍。具体科学目标如下:

(1) 提出面向商务管理的大数据可用性评估方法及相关的数据获取、融合和可视化方



法，为实现大数据分析和有效利用提供支持。

通过分析大数据的基本特征及其对商务管理的影响，从中发现大数据的可用性特征，建立面向商务管理的大数据可用性评估指标体系和可用性动态评估推理模型，在此基础上，提出空间、时间和结构一体化的数据集成框架和富媒体、流数据融合方法；提出面向商务管理的大数据可视化方法，实现相应的可视化技术。

(2) 提出语义层和表示层的大数据系统建模方法，以及面向超高维稀疏性大数据的快速分析方法，为实现大数据的有效利用提供理论与方法基础。

通过深入分析大数据环境下商务管理对数据分析与计算的新需求，从语义层和表示层提出大数据的系统建模方法；在大数据系统模型的支持下，提出面向复杂存储数据和流数据的具有线性、亚线性时间空间复杂度的高效计算方法，为大数据环境下的商务行为机理分析、企业运营优化与协调以及商务模式创新等商务管理任务提供理论和方法支撑。

(3) 提出基于网络大数据的商业情报检索、商业情报挖掘和商业趋势预测方法，为网络大数据在商业情报分析领域的应用提供理论和方法基础。

紧密结合不同层次的商务管理活动对商业情报分析方法的需求，构建全方位立体化的商业情报本体，建立大数据环境下语义关联关系表示模型和商业情报动态发现、匹配模型；以大规模语料的情感分析和基于标签的富媒体情感分析为基础，提出基于情感分析的社会媒体观点挖掘方法和基于观点挖掘的动态商业情报挖掘方法；构建客户倾向性和商业流行趋势动态演化模型，并融合其它通用方法提出面向特定商务目标的商业趋势预测方法。

(4) 构建基于大数据的用户兴趣模型，提出面向商务信息的多目标个性化推荐方法，为提升商务管理信息服务水平提供理论和方法基础。

通过分析大数据特征及其对用户兴趣形成机制的影响，分别提出面向用户个体和用户群体的兴趣建模方法和兴趣演化分析方法；针对用户多维度需求、信息多维度特征以及准确性、代表性、多样性和新颖性等多种推荐要求，提出个性化推荐的多目标优化方法；提出大数据环境下的推荐结果解释策略、个性化信息展示策略和结合企业目标的推荐优化方法，从兴趣模型构建、多目标个性化推荐以及推荐解释与实际应用等方面提升个性化信息推荐的效果。

(5) 紧密结合我国商务管理实践，构建面向商务管理的大数据研究与应用支撑平台的体系结构，开发面向商务管理的内外部数据获取系统、计算实验平台与原型系统。

通过与奇瑞汽车股份有限公司、江淮汽车股份有限公司、北京百度网讯科技有限公司、太平洋汽车网、缔元信等国内不同类型企业进行合作，充分了解大数据分析与计算方法在顾客洞察与市场营销、企业运营与商务模式创新等商务管理活动中的应用实践，从中提炼出面向商务管理的大数据分析与计算中的重要科学问题；通过典型企业商务管理的应用研究，设计出大数据支撑平台体系结构的标准规范，构建基于云架构的数据间的协同机制、基于工作流的算法间的协同机制、以及基于可视化技术的人机协同机制，实现大数据的高效并行处理环境和透明服务的大数据分析支撑环境；通过基于大数据的汽车品牌管理系统



的案例研究，对相关理论研究成果进行实践检验，并根据检验结果对相关理论与方法进行改进与提高，争取提出一套大数据环境下商务管理和应用效果的评价方法体系，开发出面向商务管理的内外部数据获取系统、计算实验平台与原型系统。

2.2 研究内容

根据大数据环境下商务管理的发展需求和《重大项目指南》的要求，在“有限目标，重点突破”原则指导下，在深入分析大数据的基本特征以及大数据在商务管理中的应用前景的基础上，按照“获取与展示（可用信息）—分析与计算（可用知识）—应用与服务（知识服务）—平台与系统（案例示范）”的思路拟定本课题的研究内容。具体分为六个专题，专题一：面向商务管理的大数据获取方法与可视化技术；专题二：面向商务管理的大数据分析与计算方法；专题三：基于网络大数据的商业情报分析方法；专题四：大数据驱动的个性化用户建模与推荐方法；专题五：面向商务管理大数据支撑平台的核心理论与关键技术；专题六：案例研究：基于大数据的汽车品牌管理系统。各专题内容的关系如图 1 所示。

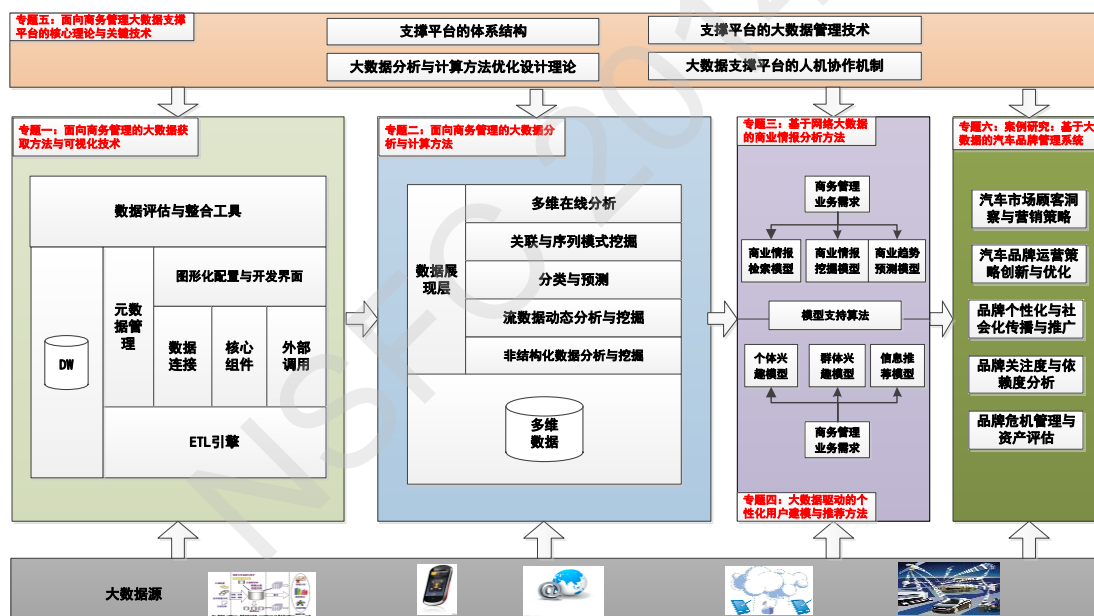


图 1 本课题各专题内容之间的关系

专题一 面向商务管理的大数据获取方法与可视化技术

商务管理的复杂性，以及大数据生成环境的开放性和不确定性，使得支持商务管理的大数据资源以多种形态碎片化地散落在不同的数据源之中，且跨平台多源化的数据往往存在着质量参差不齐、价值稀疏、满足商务管理需求的程度高低不齐等问题。同时，大数据环境下商务管理数据在数据来源、存在形式、数据结构以及时空域特性等方面存在的客观差异，尤其是富媒体的非结构化特征和流数据的跨时空和持续计算特征使得其与其它数据的融合分析存在一定的困难。此外，面向商务管理的大数据具有更强的非结构化和时空域特性，阻碍了人类的快速理解，可视化不仅能够展示大量数据中的重要信息而且能够驱动相关研究，是解决这一问题的重要工具。因此，研究面向商务管理的大数据获取方法和可视化技术已经成为大数据环境下商务管理创新亟需解决的首要问题。针对商务管理大数据



的特点,按照大数据处理的一般过程,专题一将系统地研究数据获取方法、数据可用性评价方法、数据融合方法和数据可视化技术。本专题的研究是面向大数据的商务分析与计算方法研究的基础,同时它也为其它子课题提供通用的数据获取方法和可视化技术支持。具体研究内容包括:

(1) 面向商务管理的多源大数据获取与可用性评估方法

针对具有非结构化和富媒体特征的内外商业数据,研究面向商务管理目标的大数据收集策略、高效的启发式多样性大数据提取方法(如基于众包的富媒体标注方法和针对高速流数据的实时采样方法等);针对商务管理大数据的特点,从语义层、表示层以及应用层三个视角分析大数据可用性特征;从信息源、信息体、商务需求等多个维度,研究并构建面向商务管理的多维多尺度大数据可用性评估指标体系,研究基于不确定性理论的大数据动态可用性综合评估方法;以可用性评估结果为依托,研究基于混合规则的大数据清洗方法和可用数据与不可用数据的分离技术。

(2) 面向商务管理的富媒体、流数据融合方法

系统分析跨域跨平台大数据的语义失配、语义冲突和信息不完全等特征,研究面向商务管理的数据语义标注和语义融合方法;在此基础上,分别针对富媒体数据和流数据,从空间维、时间维和结构维等三个维度,从语义层、表示层和应用层等视角,研究一体化的数据集成框架和数据融合方法,研究面向复杂分布的多源异构流数据的自动融合方法。

(3) 面向商务管理大数据分析的可视化技术

针对商务管理大数据分析的特点和特定商务管理问题的需要,重点研究平行坐标法等可视化方法的改进策略,在比较分析通用的可视化实现技术和展现形式的基础上,研究相应的选择和组合策略;结合聚类、矩阵分解等降维技术,研究超高维数据的可视化技术;研究多维时间序列数据可视化技术和基于流数据的可视化内容自动生成技术;研究基于多维叠加策略的时空信息可视化技术;研究面向商务管理数据洞察的大数据集“素描”方法等。

专题二 面向商务管理的大数据分析与计算方法

发现大数据中隐含的有用知识是大数据实践的核心价值所在。大数据的多源性、海量性、异构性以及由此产生的超高维、稀疏、非对称等数据特征使得在进行数据挖掘之前需要解决大数据的系统建模问题;存储数据是在给定时效内存储于物理设备上的数据,构建面向复杂存储数据的数据分析新方法是大数据有效应用的关键问题;数据流或特征流是大数据的普遍形态,面向单流数据和多流数据的动态数据分析方法是大数据环境下知识发现的重要研究内容。专题二是本项目研究的基础性工作,为其他课题提供数据分析方法支撑。

(1) 面向商务管理的大数据系统建模与分析方法

从语义层和表示层研究面向商务管理的大数据系统建模与分析方法。为了准确描述大数据的内容,研究基于本体知识库的大数据语义层模型,具体包括定义支持商务管理的大数据本体描述模型,建立面向领域的本体知识库,研究基于本体知识库的大数据语义统一描述模型;研究基于超图的大数据表示层模型,具体包括分析大数据的语义关联类型和概



念漂移特征,研究具有不确定性特征的大数据的语义相似性和关联性度量方法,建立基于超图的、考虑上下文制约和领域本体依赖的大数据语义关联关系表示模型。

(2) 面向商务管理存储数据的大数据分析与计算方法

针对分类、聚类、关联关系分析等商务分析任务,研究面向商务管理复杂存储数据的高效计算方法。具体内容包括:研究基于算法改进和并行式计算等策略的存储数据挖掘模型构建方法,设计面向复杂存储数据的线性、亚线性或参数估计等近似计算方法;研究面向商务管理存储数据的近似计算的样本容量优化方法、考虑代表性和多样性的数据抽样方法和具有精确度柔性的近似计算方法;研究面向数据量变和质变特征的增量式数据挖掘方法,研究基于时序特征的存储数据挖掘方法,研究数据分析算法的有效性检验方法。

(3) 面向商务管理流数据的动态分析与计算方法

研究基于界标模型、滑动窗口模型等流数据建模中的抽样理论与方法、概要数据提取方法、窗口衰减理论;针对关联关系分析、分类、聚类、离群点分析和稀有事件检测等任务,研究具有高维特征的流数据和具有倾斜分布的流数据的知识发现新方法;在流数据(数据流、特征流)概念形式化和概念发现方法等基础上,探索概念漂移特征提取方法,研究基于漂移特征的概念漂移度量方法、概念漂移预测与检测方法以及噪音建模、噪音检测和过滤方法等。

专题三 基于网络大数据的商业情报分析方法

商业情报分析是大数据分析在商务管理中最基本也是最重要的应用领域之一,各种大数据分析与发展为商业情报分析方法的改进与创新提供了机会,驱动了商业情报分析方法的发展。由于商业情报具有时效性、针对性等特点,通用的信息检索方法不能完全满足商业情报检索过程中对时效性和针对性的特殊需求,需要有针对性地对商业情报检索方法进行系统研究;不同商务管理层次对商业情报的需求不同,除了一般的情报检索方法,管理者往往希望通过各种数据挖掘和统计分析方法来发现隐匿于大数据之中的有用知识,尤其是希望能够准确预测未来可能会发生的情况,以发现新的商业机会。专题三将在网络大数据环境下,在对中文文本数据的细粒度分析和主题识别方法等关键问题进行重点研究的基础上,分别针对商业情报检索、商业情报挖掘和商业趋势预测构建系列模型和方法,以满足不同层次商务管理活动的需要。具体研究内容包括:

(1) 基于网络大数据的商业情报检索方法

在商务管理大数据环境下,针对商务管理活动对商业情报的需要,以文本数据的细粒度分析方法为基础,研究大规模语料的情感自动标注方法,研究动态语料以及领域情感词典的构建方法;基于专题二中的大数据关联关系表示模型,面向商业情报检索任务,着重研究种子资源的优选策略,研究分布式并行计算框架下基于种子资源的商业情报检索方法,研究基于语义相似性的商业情报匹配方法,建立大数据环境下商业情报资源动态发现和匹配模型。

(2) 基于网络大数据的商业情报挖掘方法



面向商务管理大数据，研究基于标签的富媒体（图片、视频或音频信息等）细粒度情感分析方法；针对商业情报的实时性要求，基于社交媒体流，研究面向商业情报挖掘的动态主题识别方法；融合专题二大数据分析 with 计算方法，结合商务管理实践的需要，着重研究面向销售业绩变化规律的大规模时序数据挖掘方法、面向客户时空分布差异的商业情报挖掘方法，以及基于流数据的动态竞争情报分析方法、企业品牌实时监测和异常状况发现方法。

（3）基于网络大数据的商业趋势预测方法

以针对文本、富媒体的细粒度分析和针对社交媒体流的动态主题识别方法为基础，研究基于多信息源的具有时空域特性的客户倾向性和商业流行趋势动态演化模型和变化趋势预测方法；融合专题二大数据分析 with 计算方法，结合商务管理实践的需要，着重研究基于大数据分析的销售量预测方法、基于商业流行趋势的产品创新机会发现方法，以及基于客户观点变化趋势的新客户群发现方法和未来市场结构预测方法。

专题四 大数据驱动的个性化用户建模与推荐方法

用户兴趣是一个多维度概念，信息需求、需求偏好、使用效用、支付意愿、认知风格等均从不同维度影响着用户的信息采纳决策，利用社交行为、搜索行为、浏览行为、评论行为、购买行为和位移行为等数据构建用户多维度兴趣模型是大数据环境下个性化推荐的基础；随着社交网络应用的兴起，群体行为成为网络用户行为的重要特征，与面向用户个体的信息推荐相比，面向用户群体的信息推荐往往具有更高的推荐效率，理解用户群体行为机理、发现目标用户群并构建群体兴趣模型，成为个性化推荐的新兴问题；个性化信息推荐过程中，基于提高代表性信息推荐效果这一目标设计的准确性指标不再是衡量推荐优劣的唯一标准，在大数据环境下解决准确性、代表性、多样性和新颖性推荐的均衡问题需要新的思路。以课题一的商务行为和课题二的市场洞察为基础，专题四的具体内容包括：

（1）大数据驱动的用户个体兴趣建模方法

研究基于大数据的用户兴趣维度识别方法，设计用户多维度兴趣图谱的表示方法；研究面向社交行为、搜索行为、浏览行为、评论行为、购买行为和位移行为等多源数据的交叉学习方法、迁移学习方法和异构网络挖掘方法，研究基于大数据的个性化推荐冷启动问题的求解策略，构建用户个体的多维度兴趣模型；分析用户兴趣的时空维特征，研究用户兴趣时空规律的分析方法，研究用户兴趣的动态预测方法，构建考虑时空维特征的用户个体兴趣演化模型。

（2）大数据驱动的用户群体兴趣建模方法

基于网络兴趣群体的数量众多、规模较小、行为稳定等特点，研究集成社会属性和兴趣属性的影响力度量方法和用户间兴趣的影响机制，分析群体集体行为(如新的习惯、流行时尚、主流观点的突然涌现)的形成机理与同步现象；研究群体用户兴趣图谱的表示方法，分析用户个体兴趣图谱和用户群体兴趣图谱的关联关系，研究基于异常点挖掘的小概率有趣模式识别方法，构建群体兴趣的挖掘模型和集结模型，保留用户群体的多样性兴趣信息，



研究不同群体之间的兴趣桥接机制和影响机制；构建群体兴趣的演化模型，研究群体兴趣及其演化的可视化方法。

(3) 考虑多样性和代表性的个性化信息推荐方法

基于用户多维度兴趣模型，利用高阶关联和迭代寻优等策略设计高维稀疏相似度问题的解决方法，设计综合考虑准确性、代表性、多样性和新颖性等推荐目标的推荐效果评价指标；研究基于用户多维度兴趣和信息多维度特征的多目标推荐方法，设计基于多推荐目标均衡的推荐列表优化方法，研究多源推荐结果的融合方法；研究用户对推荐信息信任的形成机制，基于社交网络、在线评论等开放大数据环境的推荐结果解释策略，设计基于用户认知风格的推荐信息展示策略，构建融合市场营销策略和可持续性的个性化推荐方法。

专题五 面向商务管理大数据支撑平台的核心理论与关键技术

大数据的复杂特征对信息系统的体系结构及各个层次都提出了很多非常苛刻的要求，设计面向商务管理的大数据支撑平台面临诸多挑战。本专题重点研究面向商务管理大数据平台的体系结构设计以及数据管理、计算策略、接口展示等核心理论与关键技术。

(1) 面向商务管理大数据支撑平台的体系结构

面向商务管理的大数据支撑平台从云计算和云存储的理论和模型出发，针对大规模的数据密集应用，重点研究为大数据处理提供强时效、高性能和高吞吐的体系结构设计。研究构建云计算环境下的异构数据模型，为用户提供透明、统一的异构数据集成和访问接口服务；研究大数据环境下数据资源池、知识资源池、模型资源池、方法资源池等的存储结构与构建策略；研究大数据支撑平台的集成框架与验证服务平台，实现待验证功能模块和数据库的无缝接入，实现各个子课题数据库、模型库及相关模块的动态可扩展的集成及验证操作。

(2) 面向商务管理大数据支撑平台的大数据管理技术

为满足强存储、强时效的需求，研究新的体系结构中的大数据管理技术。基于冷热数据预测，考虑商务管理视角的数据相似性对数据布局的影响，研究以存储资源占用、负载均衡、系统可靠性为目标，以存储节点的容量、流量、带宽等为约束条件的大数据存储优化方法；研究大数据在提供实时服务过程中的数据重组机制，研究基于数据冷热程度的热点数据复制转移、冷门数据合并等操作策略；研究考虑存储节点失效、低冗余的大数据可靠存储机制，构建基于贝叶斯网络的存储系统可靠性评估模型；研究大数据的时效性管理方法，基于商务管理对数据的时效要求，重点研究面向管理目标的数据时效性评估策略以及时效管理策略。

(3) 并行计算架构下基于工作流的大数据分析与计算方法优化设计理论

针对高性能并行的需求，研究新的体系结构中基于工作流的大数据分析与计算方法。结合云计算环境下大数据处理的并行计算框架，以降低任务之间的依赖性、减少不合理数据布局所引起的通信开销为目标，研究基于依赖度和并行度的复杂商务分析任务（包括关联关系分析、分类、聚类等）分解模型以及分布式并行计算框架下各子任务的计算模型；



基于求解各子任务的优先级和流程关键链，研究以优化关键链为核心的计算 workflow 设计方法；基于计算过程中数据移动和读取频率的最优化目标，以任务间的数据依赖、执行顺序为约束，研究在正常情况下以及有新任务加入、撤销、变更、结束和新计算节点的加入、退出等突发事件发生时，关键链上各项计算任务的动态调度方法；研究影响计算特性的关键要素，构造计算行为的成本函数，分析算法的时间复杂度和空间复杂度等。

（4）面向商务管理大数据支撑平台的人机协同机制

在理想的情况下，大数据的分析将不是纯计算的，而是人机交互的，如果用户无法理解分析的结果，大数据分析将失去意义，人的主观判断力与大数据分析计算方法的连接是面向商务管理大数据支撑平台面对的新挑战。围绕人机协同机制，重点研究人的参与模式、人与系统的互动模式，研究动态人机交互系统的柔性 and 可扩展性技术以及针对不同决策层次的决策服务推送机制；结合现有主流云计算平台，研究面向商务管理的可视化输出等易用性技术以及相关系统接口；研究人机交互系统计算试验场设计，并对平台中所建立的相关模型与方法进行仿真设计，利用离线分析与在线实验相结合，研究自适应反馈策略和快速迭代机制，并利用反馈机制促进人的主观决策和数据驱动决策的融合与协调。

专题六 案例研究：基于大数据的汽车品牌管理系统

在移动互联网与社会性网络快速发展的今天，企业的品牌管理面临着许多新的挑战，企业面临的客户需求多变且多样化、品牌关系维系和企业对品牌的掌控变得更加困难、品牌危机爆发具有突发性。与此同时，大数据环境却为企业品牌管理创造出了良好的机遇，企业可以快速、有效地识别目标客户群，把握消费者需求与兴趣偏好的变化；可以方便、快捷地掌握品牌的市场地位、存在的问题、未来的发展趋势；品牌关系的维护可以更具有针对性；领导者、员工、合作伙伴、消费者等都是品牌建设的积极参与者。为此，本课题以处于消费热点、市场关注度高、品牌数量规模适中、市场化程度高的汽车品牌作为案例研究对象，综合集成重大项目中各课题的研究成果，构建基于大数据的汽车品牌管理系统，从汽车市场顾客洞察、品牌运营策略创新、品牌个性化与社会化传播与推广、品牌关注度与依赖度分析、品牌危机管理与资产评估等方面，检验大数据环境下商务管理的相关理论、方法与技术。具体研究内容如下：

（1）大数据环境下汽车品牌管理系统的需求分析

分析汽车品牌管理过程中的相关任务和知识获取需求，研究大数据环境下面向汽车品牌管理的可用性数据评估、获取、聚合机制和可用数据存储的新模式，研究并建立面向汽车品牌管理的大数据系统模型；研究大数据环境下企业内部员工、供应链伙伴、消费者等相关主体在汽车品牌管理中的角色特征和关系特征，研究上述特征对汽车品牌管理的影响，分析汽车品牌管理中消费者洞察、营销和运营的新需求，并将课题二和课题三的研究成果应用于汽车市场顾客洞察与营销策略、汽车品牌运营策略创新与优化。

（2）大数据环境下品牌个性化与社会化传播与推广

针对大数据环境下汽车品牌商务管理实践，研究品牌个性化传递和社会化传播机制与



推广策略。主要研究如何利用大数据智能工具强化品牌与消费者联系,研究基于用户特征的个性化品牌推荐方法、群推荐方法,研究基于社交网络分析的社会化传播机制,构建传播效率评价体系,研究提高品牌知名度的社会化推广策略。

(3) 大数据环境下品牌关注度与依赖度分析

针对大数据环境下汽车品牌商务管理实践,研究汽车品牌关注度与依赖度分析模型。主要研究基于社交媒体挖掘的品牌关注度量化模型,揭示影响品牌依赖度的相关因素,研究品牌关注度、依赖度对顾客购买行为的影响等。

(4) 大数据环境下品牌危机管理与资产评估

针对大数据环境下汽车品牌商务管理实践,研究汽车品牌危机全周期管理与资产评估。主要研究大数据环境下汽车品牌危机预防、危机监测、危机处理和品牌重建全过程管理;研究新媒体环境下汽车品牌社会化资产评估方法。

(5) 大数据环境下汽车品牌管理系统效果评价

分析影响大数据环境下汽车品牌管理的因素,对影响因素的关系和重要性分别进行研究,在此基础上建立大数据环境下汽车品牌管理系统的评价指标体系,并结合汽车品牌管理系统结果的有效性分析,进一步完善相关研究内容,使大数据环境下面向商务管理的相关基础理论能够真正地指导管理实践。

2.3 拟解决的关键科学问题

(1) 面向商务管理的大数据可用性评估方法。支持商务管理的大数据资源以多种形态碎片化地散落在不同的数据源之中,数据往往存在着质量参差不齐、价值稀疏及噪音巨大等突出问题,构建面向商务管理的大数据可用性评估指标体系,提出面向商务管理目标的、考虑数据多源性以及数据间多维内在关联的大数据可用性综合评估方法,快速评估数据的质量及其可用性,是构建空间、时间、结构一体化的跨域跨平台数据集成框架和数据融合方法的基础。该问题的解决还为大数据系统建模和快速分析提供了可靠的数据来源,避免“垃圾进,垃圾出”,可以保障大数据环境下商务管理实践的效果,是本课题的第一个关键科学问题。

(2) 面向商务管理的超高维稀疏性大数据快速分析方法。面向多源海量数据的知识发现是数据分析的固有难题,大数据的超高维、稀疏性特征以及大数据环境下的即时服务进一步加大了数据分析的难度。通过深入分析大数据环境下商务管理创新对分类、聚类、复杂关联关系分析等大数据知识发现任务的要求,基于分布式并行计算框架,应用复杂网络理论、相似系统理论等,从上下文关联、语义关联、任务关联等不同视角,建立不同粒度的大数据关联关系网络,在此基础上,构建高效的超高维稀疏性大数据的线性或亚线性分析方法以及参数估计方法,在满足商务管理精度要求的前提下实现大数据的快速分析与计算,是面向商务管理的大数据分析的核心问题,也是大数据环境下商务管理的技术保障,因此它是本课题拟解决的关键科学问题,也是本重大项目的关键科学问题。

(3) 面向商务管理的中文主题识别方法。主题识别是面向商务管理的富媒体分析的基



础和最重要的任务，企业无论是分析客户的需求与满意度、企业自身的社会形象与品牌美誉度，还是分析竞争者的产品和服务质量等，都离不开对富媒体数据的主题识别和细粒度分析。虽然在主题识别上，已经存在一些比较有效的方法，但面向中文的、社会媒体的主题识别仍然是一个巨大的挑战，提出一种有效的面向商务管理的中文主题识别方法，既是对中文处理技术的贡献，也是本课题中涉及到的领域本体构建、商业趋势预测、客户兴趣挖掘的基础，因此它也是本课题拟解决的一个关键科学问题。

(4) 个性化信息推荐的多目标优化方法。大数据环境下的个性化推荐不仅需要均衡信息需求、用户偏好、使用效用、支付意愿等用户兴趣的不同维度，还需要考虑准确性、代表性、多样性和新颖性等推荐要求，是一个复杂的多目标决策问题。基于大数据识别用户的兴趣维度，设计用户多维度兴趣的分析方法和集成方法；基于用户多维度兴趣和信息多维度特征，构建个性化推荐的多目标优化方法，是大数据环境下面向商务管理的个性化信息推荐的核心难题，也是本课题拟解决的一个关键科学问题。

(5) 大数据支撑平台体系结构设计理论。面对大数据在规模、增长、分布、处理等方面具有的复杂特征，现有的以计算为中心的技术难以满足大数据的应用需求，因此，设计有效的大数据支撑平台体系结构是面向商务管理的大数据研究与应用的基础。平台体系结构的设计涵盖数据管理、计算策略、接口展示等方面，通过设计大数据支撑平台体系结构的标准规范，构建基于云架构的数据间的协同机制、基于工作流的算法间的协同机制、以及基于可视化技术的人机协同机制，实现大数据的高效并行处理环境和透明服务的大数据分析支撑环境。因此，大数据平台的体系结构设计是本课题拟解决的一个关键科学问题，也是本重大项目的关键科学问题。

2.4 在重大项目中所起的作用

本课题与其他课题的关系如图 2 所示。为了适应大数据环境下的商务管理创新，本课题将按照“获取与展示—分析与计算—应用与服务—平台与示范”思路系统地研究面向商务管理的大数据分析与计算的核心方法及其关键技术，开发面向商务管理研究与应用的平台，争取为本项目中的课题一“大数据环境下的商务行为机理”、课题二“大数据环境下的顾客洞察与市场营销策略”、课题三“大数据环境下的运营策略优化与协调”以及课题四“基于大数据的商务智能与模式创新”提供数据获取、分析与计算方法及其关键技术方面的支持，也可以为其他课题的模型、方法验证提供平台支持，对支撑“大数据环境下的商务管理”的协同研究具有重要作用。同时，大数据环境下的商务行为机理、大数据环境下的顾客洞察与市场营销策略、大数据环境下的运营策略优化与协调等课题的研究成果对本课题的模型构建、算法设计，以及基于网络大数据的商业情报分析、大数据驱动的网络用户建模与个性化推荐方法等专题的研究具有指导作用。

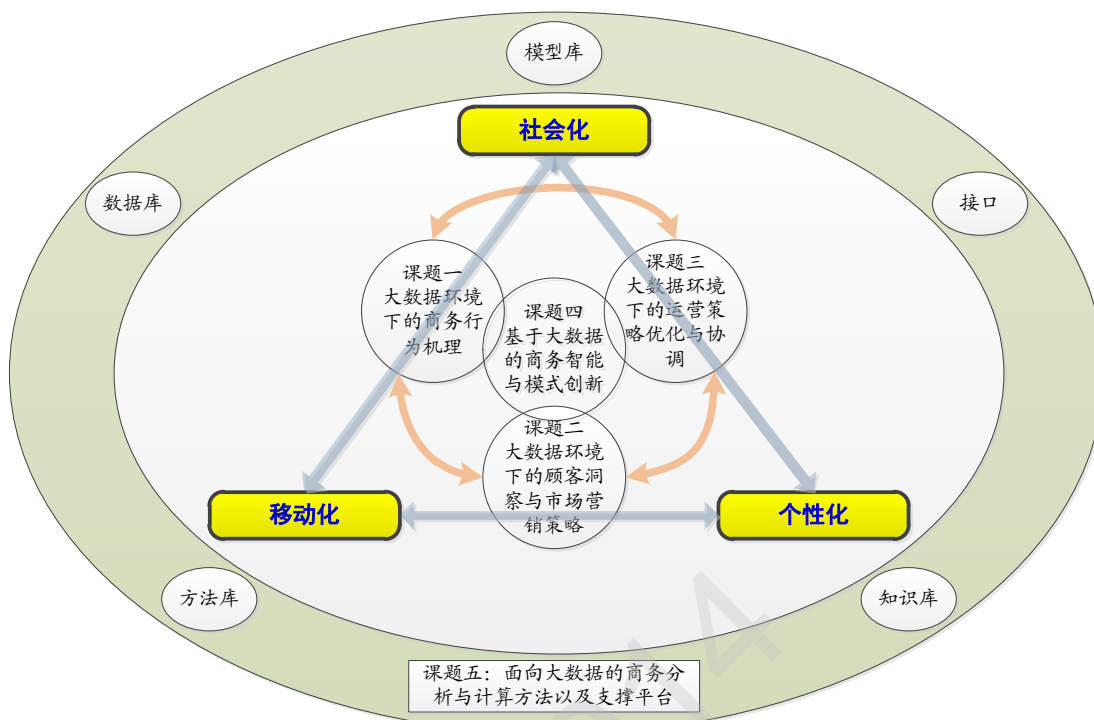


图2 本课题与其它课题间的关系

3、课题的研究特色与创新之处。

通过分析面向大数据的商务分析与计算方法及支撑平台的国内外研究现状，以企业的商务管理实践为引导，科学系统地设计本课题的研究内容，凝练五个亟待攻克的关键科学问题，并以数据密集型研究范式寻求“事实”作为研究/思维的逻辑起点，采取理论研究、调查研究、模拟仿真和实例研究相互促进的方法开展研究。通过深入的研究工作，预期的创新点如下：

(1) 提出大数据可用性评估模型及评估方法。该方法基于证据推理理论，针对商务大数据多源异构、质量参差不齐、碎片化等特点，从信息源、信息体、商务管理需求等多个维度揭示大数据在语义层、表示层和应用层的可用性特征，构建面向商务管理的多维多尺度大数据可用性评估指标体系和综合评估方法，在此基础上提出“空间、时间、结构”一体化的数据融合与集成框架，从而实现在建立大数据可用性评估理论上的创新。

(2) 提出面向超高维稀疏性商务管理大数据的计算模型与快速分析方法。该模型基于诸如深度神经网络、多尺度无监督学习、最大信息系数、流形学习、分形理论、小波分析等先进的机器学习、统计学理论，针对商务大数据中广泛存在的超高维、稀疏性等问题，提出有效的抽样方法、数据压缩方法、主数据分析方法以及相应的统计检验方法，在保证商务分析的精度要求下，构造商务管理大数据的线性、亚线性或参数估计等快速分析方法，从而实现大数据环境下数据分析与计算理论的创新。

(3) 提出基于标签自动标注和情感分析方法的中文主题识别方法。该方法基于领域本体、标签语料库、情感语料库，针对中文富媒体语料，构建情感符号的自动标注方法，在



此基础上,设计出基于标签的情感分析方法和基于文本情感分析方法的中文主题识别方法,实现商业情报分析方法上的创新。

(4) 提出考虑多样性、代表性等多目标优化的个性化信息推荐方法。该方法将个性化信息推荐看作是既要考虑信息需求、用户偏好、使用效用、支付意愿等用户兴趣维度,又要考虑准确性、代表性、多样性和新颖性等推荐维度的多目标决策问题,基于此,提出大数据环境下个性化推荐的相似性度量方法、匹配度计算方法和推荐效果评价指标以及个性化推荐的多目标优化方法,并设计个性化推荐的解释策略、展示策略和营销策略,从而实现大数据环境下个性化信息推荐理论的创新。

(5) 提出以数据间的协同、计算间的协同、人机间的协同为核心的面向商务管理的大数据支撑平台体系结构,并结合我国汽车制造企业在大数据环境下的商务管理实践,开发面向商务管理的大数据支撑平台系统原型。以奇瑞汽车股份公司、江淮汽车股份公司等企业的基于大数据的汽车品牌管理为应用背景,并结合百度、太平洋汽车网、缔元信等互联网企业,提炼并解决面向大数据的商务分析与计算方法以及支撑平台的关键科学问题,形成一套面向商务管理的大数据分析与计算理论、方法和技术,构建以“基于云架构的数据间的协同、基于工作流的算法间的协同、基于可视化技术的人机间的协同”为核心的面向商务管理的大数据支撑平台体系结构,在此基础上开发大数据环境下面向商务管理的支撑平台系统原型,从而实现对大数据环境下商务管理理论方法的检验和分析计算工具的创新。

4、拟采取的研究方案及可行性分析。(包括有关方法、技术路线、实验手段、关键技术等说明)

4.1 研究方案和技术路线

(1) 总体技术路线

课题总体技术路线如图3所示。有关技术路线说明如下:

1) 本课题拟首先通过文献调查对国内外有关面向商务管理的大数据分析与计算的核心方法和关键技术与应用成果进行分析总结,同时深入安徽江淮汽车股份有限公司、奇瑞汽车股份有限公司、北京百度网讯科技有限公司、太平洋汽车网等相关企业进行实际调研,了解大数据在商务管理中的实践经验和创新需求,再结合申请人和项目组成员前期的研究成果,完成课题的前期调研和分析工作。在分析并深入理解前期调研成果的基础上,明确大数据环境下商务分析与计算的相关需求,对课题总体研究思路及方案进行系统、合理的规划并形成初稿,结合所形成的研究方案,再次进行文献分析与企业调研,最终确定课题的总体研究思路及方案。

2) 在研究过程中,针对关键科学问题,采用理论研究、调查研究、模拟仿真与案例研究相互促进的方法开展具体的研究工作。在理论研究过程中,主要研究面向商务管理的大数据可用性评估理论、面向商务管理的大数据系统建模理论和面向商务管理的超高维稀疏

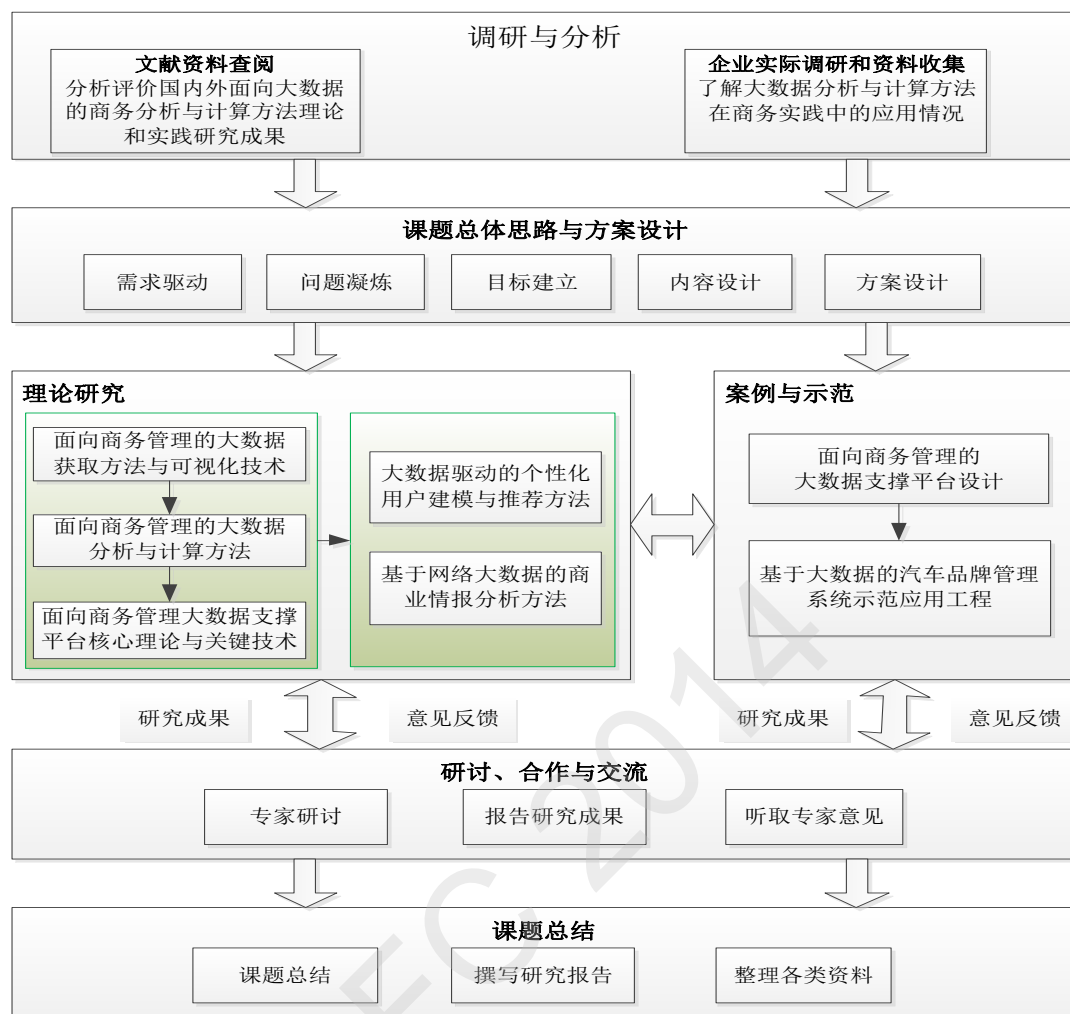


图3 课题总体路线

性大数据快速分析方法、面向商务管理的中文主题识别方法、个性化信息推荐的多目标优化方法、大数据支撑平台体系结构设计理论，为开展仿真研究和案例与示范研究奠定理论基础。在案例与示范研究过程中，以基于大数据的汽车品牌管理系统为实际应用背景，研究具有针对性的理论问题，开发面向汽车品牌管理系统的实时动态知识服务系统原型，并结合案例与示范研究过程中遇到的问题，进一步拓宽研究思路，扩充理论研究内容。

3) 结合课题研究中的关键科学问题，积极参与国内外学术交流，并在适当的时候组织召开专题学术研讨会，向相关领域专家和企业人员报告课题研究的阶段性成果，听取他们的建议和意见，努力攻克课题中的关键科学问题。最后对课题研究进行总结，撰写课题的研究报告并整理研究过程中的各类资料和成果。

下面针对每个专题，详细展开其研究的具体路线。

(2) 专题一 “面向商务管理的大数据获取方法与可视化技术”技术路线

“面向商务管理的大数据获取方法与可视化技术”的技术路线如图4所示。拟采取的研究方案和关键技术路线具体阐述如下：

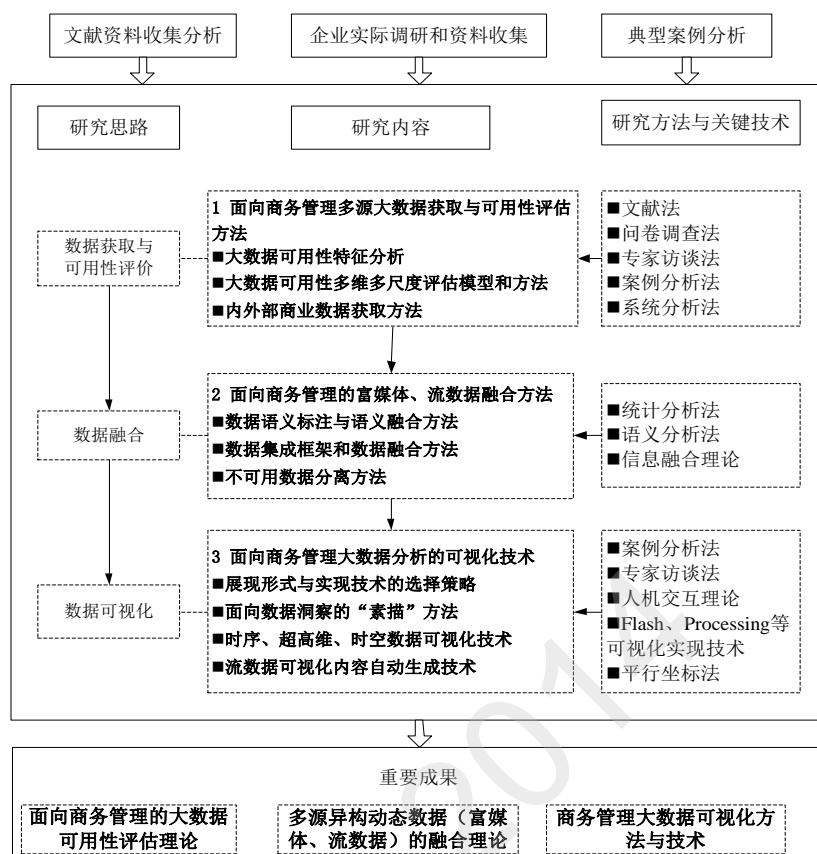


图4 专题一“面向商务管理的大数据获取方法与可视化技术”技术路线

1) 从理论研究和实际应用两个方面收集关于大数据特征及可用性特征的相关文献资料,应用文献研究法从信息源、信息体、管理决策需求等多个维度,分析大数据可用性的内涵和外延,从语义层、表示层和应用层分析大数据可用性特征以及可用性的影响要素;应用问卷调查、相关性分析、系统分析等方法,建立面向管理决策的大数据可用性评估指标体系;应用案例分析法和相关评估理论,研究基于多维属性的大数据可用性综合评估方法,应用证据理论等多准则决策方法,建立大数据可用性的评估模型;通过文献资料收集和企业实际调研两种手段,明确面向商务管理的大数据收集策略,研究高效的启发式多样性大数据提取方法。

2) 应用统计分析和语义分析法分析跨域跨平台大数据的语义失配、语义冲突和信息不完全等特征;采用关联分析方法和语义分析法研究多源异构大数据之间的内在联系;分别针对富媒体数据和流数据,应用证据理论等信息融合理论,从空间维、时间维和结构维等不同维度研究一体化的数据集成框架和数据融合方法;应用哈希技术研究可用数据与不可用数据的分离技术。

3) 通过专家访谈和案例分析法明确常见的商务管理大数据可视化领域的特定问题和场景,提出可视化技术(如 Processing、Flash、OpenCV 等)与展现形式(如堆积图、热力图、网络图等)的选择策略。结合人机交互理论和大数据分析技术,研究大数据的“素描”方法和流数据的可视化内容自动生成技术等;结合人机交互理论和地理信息可视化方法,研究基于多维叠加策略的时空信息可视化技术;改进平行坐标法等多维可视化技术,结合聚类、



矩阵分解等降维技术，研究多维时间序列数据、超高维商务数据的可视化技术。

(3) 专题二 “面向商务管理的大数据分析与计算方法”技术路线

“面向商务管理的大数据分析与计算方法”的技术路线如图 5 所示。拟采用的研究方案和技术路线具体阐述如下：

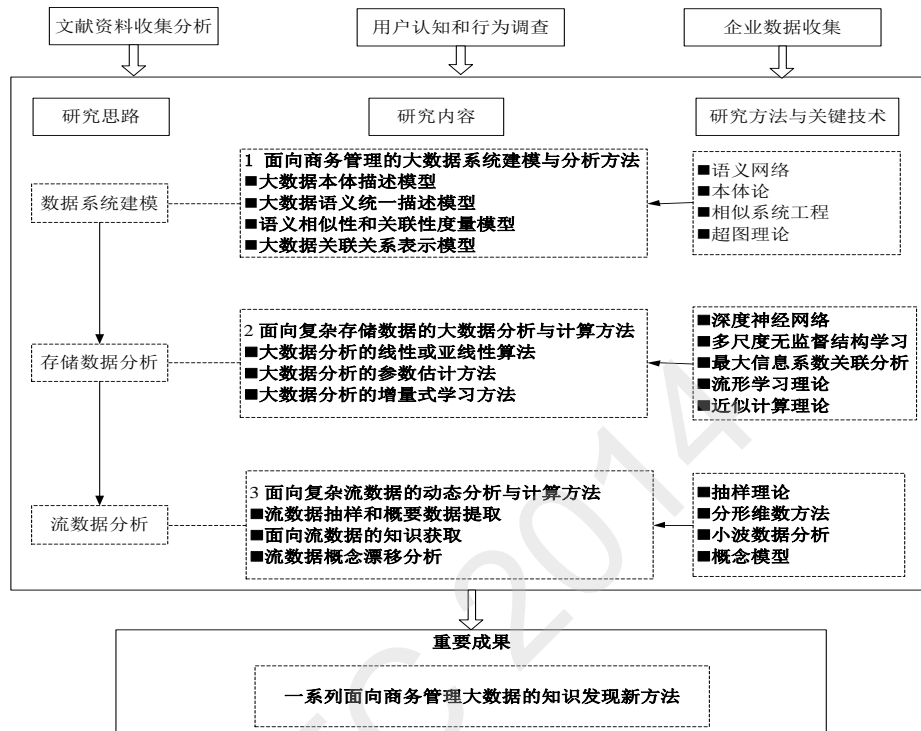


图 5 专题二 “面向商务管理的大数据分析与计算方法”技术路线

1) 应用语义分析法，研究大数据语义层和表示层的内在联系，采取形式化描述的方法对大数据的属性、结构、交互、联合、控制等基本元素、基本操作和控制结构进行定义，借助本体论和语义网，建立大数据统一的语义描述模型；应用相似系统工程的相关理论和方法以及关联关系分析，研究不同粒度数据之间的语义相似性和关联关系度量方法，应用超图理论和复杂网络理论，表示大数据复杂的语义关联关系模型。

2) 基于文献分析法深入剖析利用现有数据分析方法挖掘大数据所面临的困难，结合对大数据多源海量性、超高维、稀疏性等特征的分析，设计基于流形学习的超高维大数据的降维方法和基于多尺度无监督结构学习的稀疏性大数据表示结构分析方法，从基准点选择、规则生成、剪枝策略以及增量学习策略、近似学习策略等方面对传统的 K-means、Apriori、C4.5 等数据挖掘算法进行改进，设计与大数据环境下相适应的深度学习方法和最大信息系数的关联分析等方法，基于数据压缩、主数据分析和并行计算等策略探索上述方法的线性/亚线性复杂度设计思路，实现大数据环境下的快速降维、分类、聚类、关联和时序分析等任务。

3) 通过实证分析和典型案例分析方法，总结大数据环境下多源、多维流数据的特征，利用时间戳、滑动窗口等技术设计商务管理流数据抽样模型；改进时间窗口、决策树、时序频繁模式挖掘等经典流数据挖掘方法，创新分形数据挖掘、小波数据分析等方法，设计



面向商务管理流数据的知识获取方法和概念漂移分析方法；应用复杂性理论、系统优化方法等分析、优化所设计算法的性能，并应用仿真实验验证算法的有效性。

(4) 专题三 “基于网络大数据的商业情报分析方法”技术路线

“基于网络大数据的商业情报分析方法”的技术路线如图 6 所示。拟采取的研究方案和关键技术路线具体阐述如下：

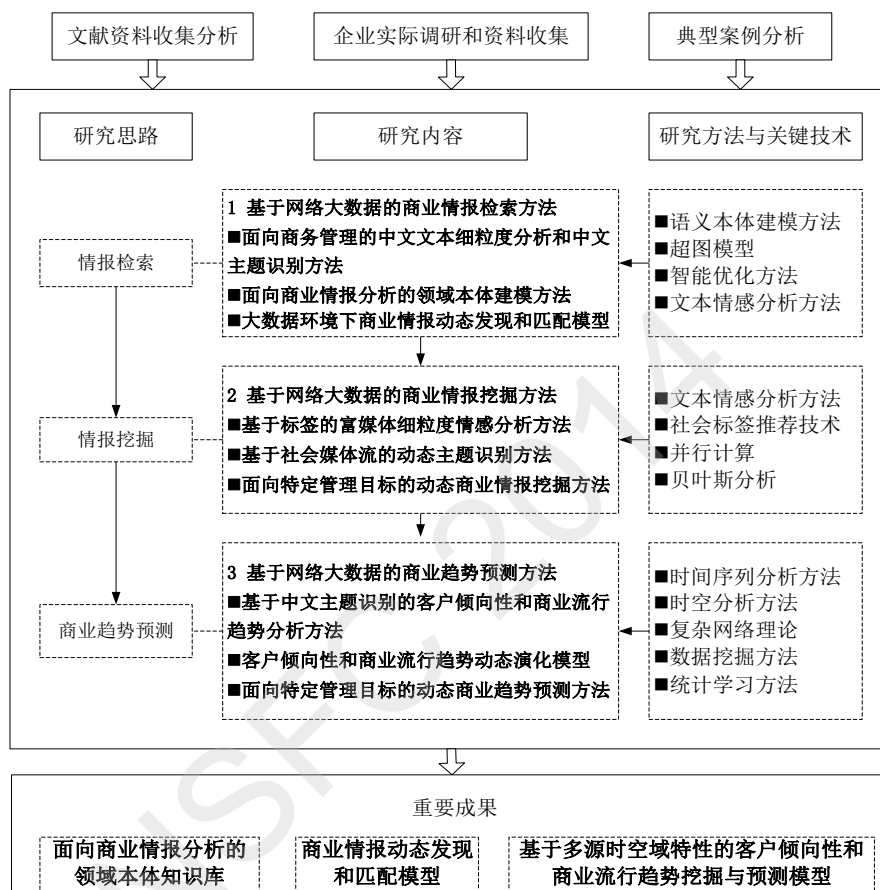


图 6 专题三“基于网络大数据的商业情报分析方法”技术路线

1) 通过对企业实际调研，选取典型商务管理领域的商业情报分析应用场景，分析大数据环境下企业商业情报检索的需求类型；利用文献分析法以及点互信息理论、本体论以及矩阵分解等理论构建针对商业情报分析领域的情感分析词典；针对商务管理活动对商业情报的实际需要，应用语义本体建模方法和超图模型，构建面向商业情报分析的领域本体，建立面向商业情报分析的语义关联关系表示模型；应用智能优化方法，设计基于语义相似性的商业情报匹配方法，建立大数据环境下商业情报动态发现和匹配模型。

2) 综合利用土耳其机器人和情感符号自动标注等方法构建图片、动画、声音、视频等富媒体的标签语料库；采用并行计算框架，扩展现有情感分析算法，设计大数据环境下基于标签的富媒体情感分析方法；针对流数据形式的多源异构社交媒体数据，应用大规模并行计算方法和基于文本情感分析的主题（观点）识别方法，设计面向商业情报挖掘的基于社交媒体流的动态主题识别方法；利用动态分层模型和时序贝叶斯方法挖掘面向特定管理目标的动态商业情报。



3) 以针对文本、富媒体的细粒度情感分析和针对社会媒体流的动态主题识别方法为基础, 综合应用时间序列分析方法、时空分析方法和统计学习模型等数据分析方法, 预测社会媒体主题识别的客户倾向性和商业流行趋势; 依据数据的时间、地理特性, 综合运用约束波兹曼模型、递归神经网络以及隐马尔科夫模型等方法挖掘隐藏在大数据中的商业情报分析的重要特征, 结合人类情绪季节性周期变化等心理学和行为学相关理论, 构建基于多信息源的具有时空域特性的客户倾向性和商业流行趋势动态演化模型, 并使用模拟或真实数据对模型进行验证; 应用动态演化模型提出客户倾向性和商业流行趋势预测方法, 并结合课题应用示范的具体场景, 设计面向特定目标的商业趋势预测模型和方法。

(5) 专题四 “大数据驱动的个性化用户建模与推荐方法”技术路线

“大数据驱动的个性化用户建模与推荐方法”的技术路线如图 7 所示。拟采用的研究方案和技术路线具体阐述如下:

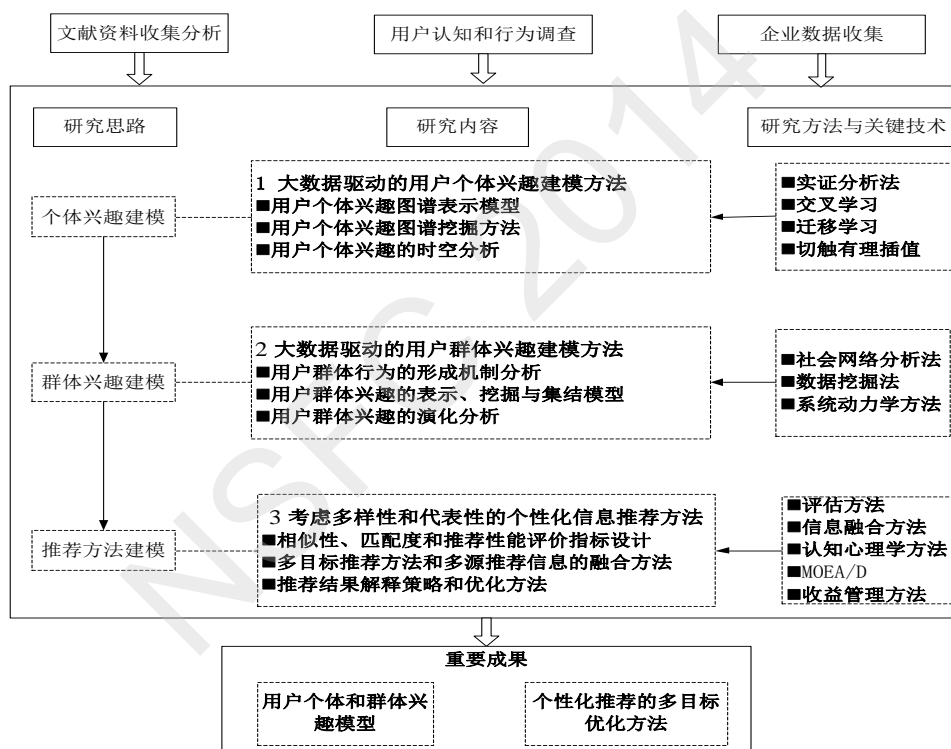


图 7 专题四 “大数据驱动的个性化用户建模与推荐方法”技术路线

1) 通过问卷调查和实证分析研究用户个体行为规律及其分布特征, 利用数据挖掘、跨平台数据交叉学习和迁移学习等手段, 分析用户个体行为特征与用户兴趣的关系, 构建基于全网数据的用户个体兴趣图谱; 利用时序数据挖掘、切触有理插值等方法构建用户兴趣的时空模型, 对用户兴趣进行动态预测。

2) 通过问卷调查、案例分析和实证研究方法, 分析用户群体的形成机制和群体行为的形成机理, 在现有影响力度量方法的基础上, 引入用户节点的社会属性和兴趣属性, 设计用户影响力的度量方法, 分析用户兴趣的相互影响机制; 利用文本分析、关联分析等数据挖掘方法挖掘群体兴趣, 从多样性数据保留、多样性兴趣挖掘以及多样性列表生成等方面设计多样性推荐列表的生成方法, 基于社会网络分析方法研究群体兴趣的桥接机制和影响



机制；利用系统动力学方法研究群体兴趣的演化规律。

3) 利用高序关联、迭代寻优等方法解决推荐过程中的稀疏相似度问题，分析现有的个性化推荐效果评价指标，设计综合考虑用户信息需求、需求偏好、使用效用、支付意愿、认知风格等兴趣维度以及准确性、代表性、多样性、新颖性等推荐目标的推荐效果评价方法；结合多目标决策方法和基于分解的多目标演化方法（MOEA/D）设计面向用户多维度兴趣、信息多维度特征和多维度推荐目标的个性化推荐方法，设计基于证据理论的多源推荐信息融合方法；利用关联分析、心理学问卷调研和企业收益管理等方法，并结合社交网络、在线评论等数据源，设计个性化推荐的解释策略、展示策略和考虑企业收益的推荐策略，优化个性化推荐的效果。

（6）专题五 “面向商务管理大数据支撑平台的核心理论与关键技术”技术路线

“面向商务管理大数据支撑平台的核心理论与关键技术”的技术路线如图 8 所示。拟采用的研究方案和技术路线具体阐述如下：

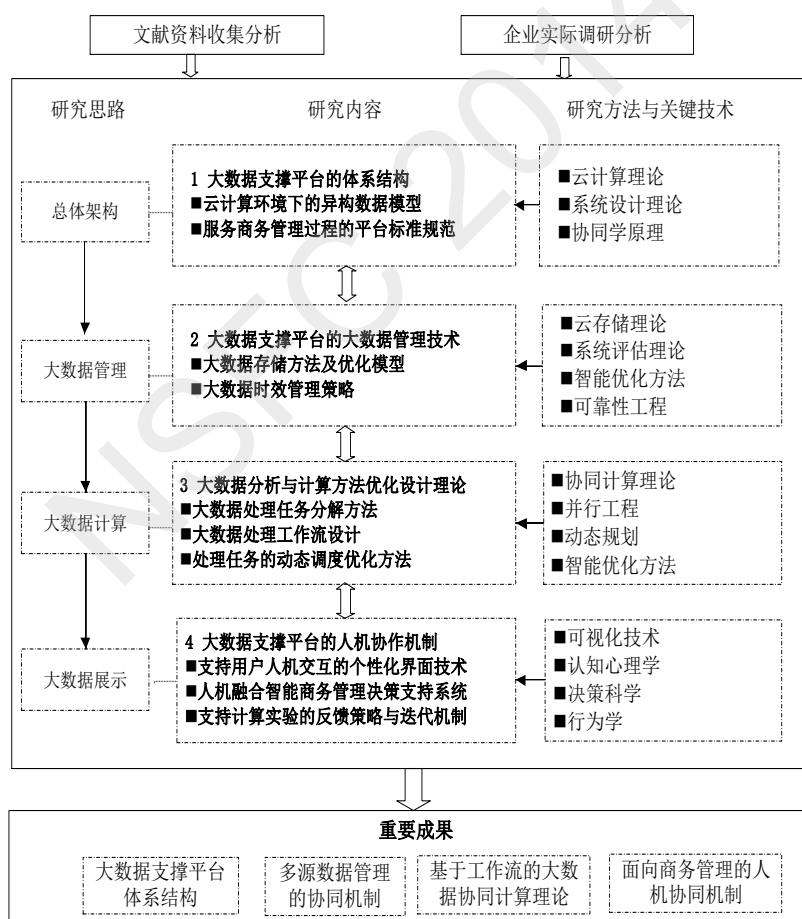


图 8 专题五 “面向商务管理大数据支撑平台的核心理论与关键技术”技术路线

1) 应用文献研究法分析、发现现有主流云计算平台的架构形式与特征，依据商务管理中不同领域、不同层次、不同需求的大数据处理特征，构建大数据支撑服务平台的体系结构，在平台上实现内外部数据获取、存储、管理、分析和决策等资源和服务的共享、交易与协作。



2) 应用文献研究法分析、发现现有大数据存储模型的特征,依据云存储理论,实现数据间的协同管理。利用相似性度量方法分析数据相关性,并利用哈希函数研究数据分块策略,运用多目标规划等方法构建存储优化模型,应用智能优化方法研究模型的求解策略;针对实时服务过程中的数据重组织研究,通过引入矩阵理论构造冷热数据的聚类矩阵,采用修正的哈希函数加大冷门数据的分配权重,构造热门数据的复制转移与冷门数据的合并操作;在存储系统的可靠性评估方面,应用可靠性工程,建立基于贝叶斯网络的存储系统可靠性评估模型,并利用仿真技术验证模型的有效性。依据实体的时间戳、实体记录或者实体属性中部分隐藏的时序信息,应用价值评估模型分析不同时序上的数据价值,并据此给出时效性管理策略。

3) 应用文献研究法分析大数据处理的主要任务类型,在此基础上应用协同计算理论研究算法间的协同任务。采用并行工程、相似系统工程、动态规划等理论,构造基于任务优先级的大数据处理任务的工作流优化模型,应用关键链理论分析工作流中的关键任务节点,研究构造出基于并行度的任务调度优化模型,应用智能优化方法研究模型的求解方法。

4) 应用文献研究法分析大数据环境下人机交互协作机制,应用可视化技术研究支持用户个性化人机交互界面设计;结合认知心理学与决策科学相关理论,分析用户主观决策与数据驱动决策的协调关系,研究构建人机融合的智能商务管理分析系统;利用计算平台设计理论与方法,研究大数据支撑平台支持商务管理研究应用的结果反馈策略与迭代机制。

(7) 专题六 “案例研究:基于大数据的汽车品牌管理系统”技术路线

“案例研究:基于大数据的汽车品牌管理系统”技术路线如图9所示。拟采用的研究方案和技术路线具体阐述如下:

1) 首先根据实验方案设计的内容构建实验云平台和数据采集系统。通过对实体汽车企业(奇瑞、江淮等)和有关汽车行业的互联网企业(百度、太平洋汽车网、缔元信等)实时数据收集,以及针对互联网上的主流社交媒体以及企业运营的售后服务平台等的及时跟踪,结合企业内部的管理数据,形成面向汽车品牌管理的大数据环境。

2) 针对汽车品牌商务管理实践,利用本课题以及本重大项目其他课题的研究成果,对汽车品牌管理中的数据获取和数据分析等问题进行研究,验证相关研究成果的有效性。在此基础上,结合针对奇瑞公司、江淮汽车等的现场调研,获得大数据环境下汽车品牌管理不同阶段的新需求,研究大数据环境下汽车品牌管理中消费者洞察、运营与营销的新需求。

3) 采用理论研究和案例研究相结合的方式研究大数据环境下汽车品牌管理过程。基于个性化推荐方法、社交网络分析方法、社交媒体挖掘方法等研究成果,研究品牌个性化传递方法和社会化传播机制;基于语义分析方法、情感分析方法、科学计量方法等研究成果,构建汽车品牌关注度量化模型,研究品牌依赖度影响因素;基于文本挖掘方法和社交媒体挖掘方法等研究成果,研究大数据环境下汽车品牌危机全周期管理和资产评估效果。

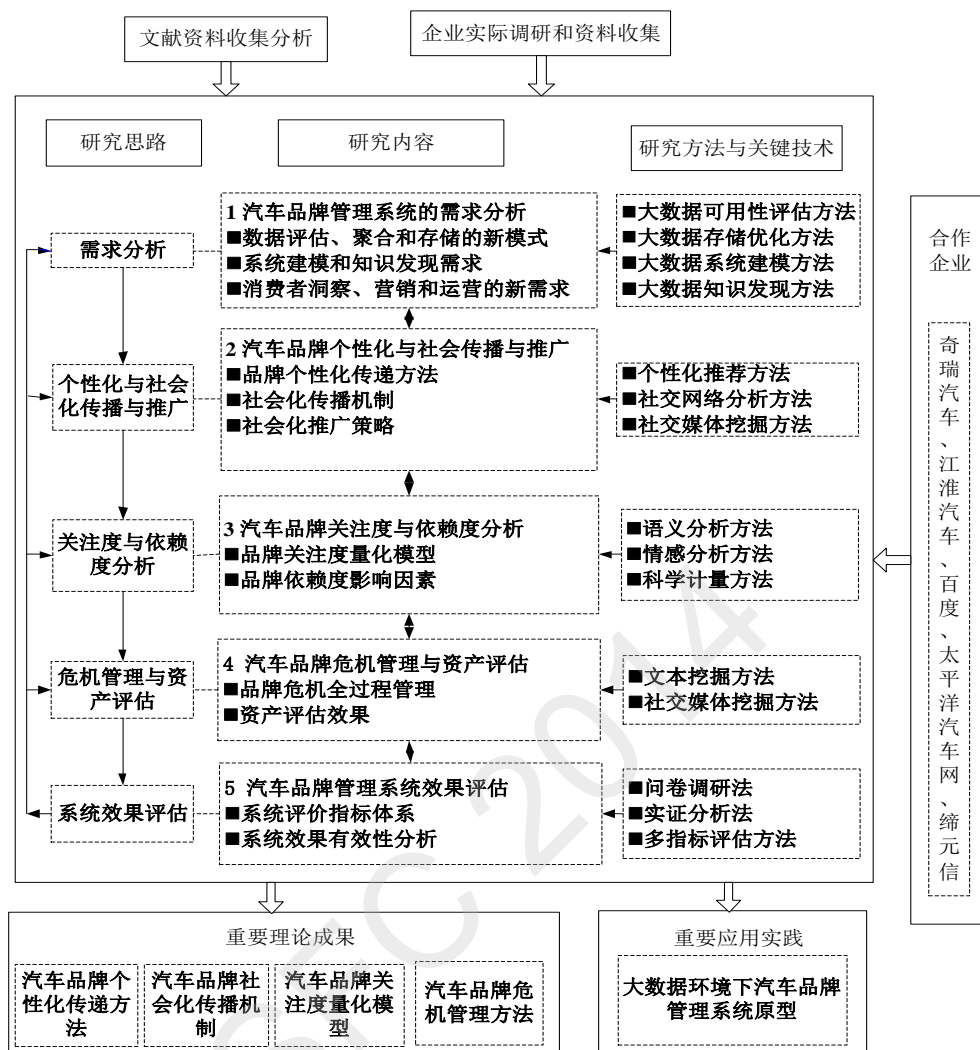


图 9 专题六“案例研究：基于大数据的汽车品牌管理系统”技术路线

4) 跟踪奇瑞汽车、江淮汽车品牌管理的实际案例，利用问卷调查法和实证分析法获得品牌管理系统效果的评价指标体系，研究基于多指标评估理论的商务管理有效性评估模型，基于评估模型进一步完善相关研究理论、方法和技术成果。

4.2 实验方案设计

本课题将以大数据驱动的汽车品牌管理系统为应用背景，建立小型云计算实验平台，支持大数据收集与评估、分析与计算以及营销洞察、运营优化、商业情报分析和个性化信息服务等模型与方法的验证，支持面向商务管理的内外部数据获取系统以及原型系统开发，保证大数据环境下商务管理的理论研究与案例研究的顺利实施，验证本项目研究成果的有效性、合理性，提高成果的可集成性与展示度。

平台拟利用“智能决策与信息系统技术”国家与地方联合工程研究中心和“过程优化与智能决策”教育部重点实验室现有的基础条件，建立包含 2 台小型机、30 台 PC 服务器和 200 台左右的 PC 机以及相关网络设备组成的硬件环境，采用 Hadoop、Spark 等软件系统作为分布式计算平台，实验平台总体架构如图 10，硬件逻辑架构如图 11，软件逻辑架构如图 12。

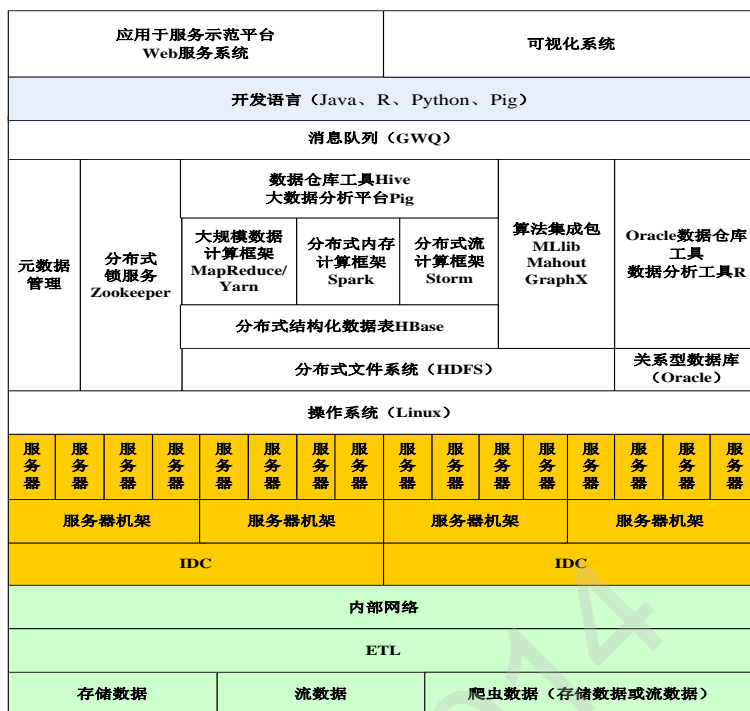


图 10 实验云平台总体架构图

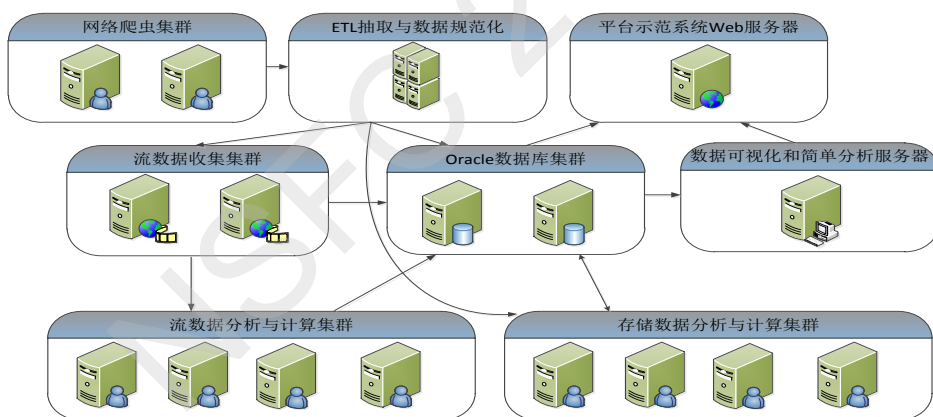


图 11 实验云平台硬件逻辑架构图

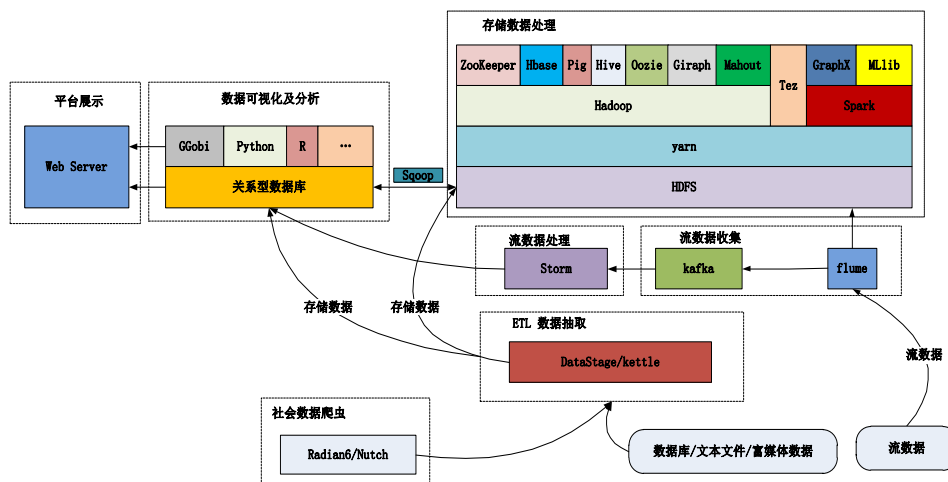


图 12 实验云平台软件逻辑架构图



1) 多源数据获取。面向汽车品牌监控与管理的大数据主要包括企业内部的业务、生产、销售等数据, 互联网上的社交媒体数据等。不同的数据会通过不同的处理进入大数据存储系统, 其中互联网上的社交媒体数据将通过 Radian6、爬虫程序 Nutch 等获取, 并与企业的已有数据一起通过 ETL 工具 Kettle 进入存储系统; 实时流数据将通过抽样进入存储系统或者通过分析将结果存入存储系统。

2) 大数据存储系统与计算平台。大数据存储系统分为两种类型: 一是分布式文件系统(如 HDFS) 以及在其之上的分布式数据库系统(如 HBase), 并在此存储系统上部署适合分布式计算的计算框架, 包括分布式计算框架(Hadoop)、内存计算框架(Spark)和流数据计算框架(Storm), 分布式的锁服务将采用 Zookeeper; 二是传统的关系型数据库系统, 如 Oracle(可支持数据仓库)。HBase 与 Oracle 间通过 Sqoop 实现数据转换。

3) 大数据分析与计算工具。除了 Hadoop 基本的分布式计算框架外, 为了使实验平台提供良好的数据分析环境, 本实验方案将在分布式存储系统与计算平台上部署分布式数据仓库工具(Hive), 并使用 Mahout、MLlib、GraphX 等工具集成经典的数据挖掘算法, 以满足后期的数据分析要求和算法实验环境, 同时提供如 Java、Python 等高级编程语言环境。

4) 可视化以及服务接口。可视化系统包括数据的可视化和大数据分析结果的可视化, 主要开发工具有 HTML5 Canvas、JavaScript、Flash, 或者使用一些开源的 JS 包, 如 iCharts、Fusion Charts Suit XT; 采用 SOA 的架构为用户提供服务, 包括原型系统的使用者或模型验证的研究者。Web 系统的服务器软件使用 Apache 的开源软件 Tomcat。

4.3 可行性分析

(1) 研究内容清楚, 研究目标明确, 技术路线合理, 研究方法可行。项目组做了大量的前期准备工作, 查阅了大量国内外的相关研究文献, 规划了比较清晰的研究框架和研究内容, 并提炼了其中的关键科学问题, 确定了比较明确的研究目标。在明确研究内容和研究目标以及大量文献检索的基础上, 综合应用系统评估理论、系统优化理论、统计学、人工智能、不确定性理论、博弈论等理论方法和技术解决关键科学问题; 在研究过程中, 采用理论研究、调查研究、模拟仿真、案例与示范研究相互促进的方法推动项目研究的深入开展, 可有效地保障本课题研究内容和研究目标的实现。

(2) 研究基础扎实。申请人及项目组成员多年来一直从事电子商务、数据挖掘、个性化推荐系统以及商务智能等方面的研究与开发工作, 取得了一批得到国内外同行和有关企业较高评价的研究成果, 在国内外顶级期刊上发表了若干高水平的学术论文、出版了多部研究专著并且承担了多项与本课题相关的国家自然科学基金项目, 积累了丰富的研究经验。

(3) 团队结构合理。项目团队由具有较高学术水平的学术带头人和有着丰富研究经历、学科组成以及年龄结构合理的研究人员组成, 包括多年来一直从事相关领域研究工作的中国工程院院士、教授、副教授, 以及对本领域拥有强烈的研究兴趣, 并具有扎实基础的青年教师、博士生以及硕士生共 35 人左右, 研究实力较强; 研究团队中的绝大多数教师均有国际学术访问和交流的背景和经验, 并与美国匹兹堡大学、佛罗里达州立大学、英国曼彻



斯特大学、布鲁内尔大学、澳大利亚斯韦本科技大学、加拿大多伦多大学、麦克马斯特大学以及香港城市大学等多所国际一流知名大学建立了长期稳定的国际学术交流与合作关系。另外，项目团队成员学科背景集成了工商管理、管理科学、计算机科学以及电子商务与信息系统等领域，具有良好的交叉学科综合研究优势，并且在系统建模、案例分析、统计分析以及数理分析等研究技能方面，具有全面而扎实的基础。

(4) 具有良好的学科支持和科研条件。申请人所在的合肥工业大学管理学院拥有“管理科学与工程”国家级一级重点学科和“工商管理”一级学科博士授权点，并拥有“智能决策与信息系统技术”国家-地方联合工程研究中心、“过程优化与智能决策”教育部重点实验室、“智能决策与信息系统技术”教育部工程研究中心、“知识经济与企业管理创新”安徽省人文社会科学重点研究基地、“决策科学与信息系统技术”教育部创新团队，具有雄厚的科研实力和良好的工作条件，这是本项目进行研究的坚实后盾。

(5) 广泛的企业合作基础。项目组在以往的研究过程中与芜湖奇瑞、江淮汽车、百度公司、阿里巴巴等企业建立了密切的合作关系，为研究工作的实地调研提供了保障。

以上分析表明，本项目团队和合肥工业大学完全有足够的能力和条件完成该项课题的研究。

5、国际合作与交流计划安排情况。

本项目在执行期间拟开展以下国际合作与交流：

(1) 定期召开远程视频会议。考虑到有海外成员参加本项目的研究，拟每月召开一次多方远程视频讨论会。

(2) 派出研究人员建立海外研究合作关系。项目组绝大多数成员都有海外研学经历，与美国匹兹堡大学、佛罗里达州立大学、英国曼彻斯特大学、布鲁内尔大学、澳大利亚斯韦本科技大学、加拿大多伦多大学、麦克马斯特大学以及香港城市大学等高校建立了良好的合作关系，在项目研究期内，项目组成员赴国外参与短期国际学术交流合作 5 人次以上，派出 2 名以上博士生前往国外知名大学进行联合培养学习。

(3) 积极主办和参与国际学术会议。在项目执行期间，拟组织举办 2 次国际会议和 4-6 次专题研讨会，同时每年派出团队成员参加相关国际重要学术会议，针对本项目的学术思想和部分研究进展，与国内外同行进行交流。

(4) 邀请国外新学者加入研究。拟邀请除课题组成员以外的国外同行学者来华开展合作交流与研究，并将针对本项目研究方向建立新的国际合作研究关系。



6、年度研究计划及预期研究结果。

6.1 年度研究计划

表 1 年度研究计划表

序号	研究内容	2015 年				2016 年				2017 年				2018 年				2019 年			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	总体思路设计																				
2	面向商务管理的内外部数据获取系统、 计算实验平台以及原型系统建设																				
3	大数据在商务管理中的应用前景分析																				
4	多源大数据获取与可用性评估方法																				
5	富媒体、流数据融合方法																				
6	大数据系统建模与分析方法																				
7	复杂存储大数据的分析与计算方法																				
8	复杂流数据的动态分析与计算方法																				
9	基于网络大数据的商业情报检索方法																				
10	基于网络大数据的商业情报挖掘方法																				
11	基于网络大数据的商业趋势预测方法																				
12	大数据驱动的用户个体兴趣建模方法																				
13	大数据驱动的用户群体兴趣建模方法																				
14	考虑多样性和代表性的个性化信息推荐方法																				
15	大数据支撑平台的体系结构																				
16	大数据支撑平台的数据管理技术																				
17	大数据支撑平台的人机协作机制																				
18	基于工作流的大数据分析与计算方法 优化设计理论																				
19	案例研究：基于大数据的汽车品牌管理 系统研究与开发																				
20	总结报告撰写																				



6.2 预期研究成果

本课题的主要预期研究成果将对大数据环境下的商务管理相关领域做出重要创新性贡献，总体上达到国际前沿水平，并在部分方向上居于国际领先地位。本课题将围绕面向大数据的商务分析与计算方法以及支撑平台开展研究，提出面向商务管理的大数据可用性评估理论、分析与计算的核心方法，解决其中的关键技术，提出基于网络大数据的商业情报分析方法和大数据驱动的个性化信息推荐方法，构建面向商务管理的内外部数据获取系统、计算实验平台以及原型系统。

同时，研究成果将陆续以论文、专著或专利形式发表。计划在国内外发表一批高水平论文，包括主流国际期刊（SCI/SSCI）论文数不少于 30 篇，其中顶尖国际期刊（即管理、数据分析类顶尖刊物：UT Dallas 24 种期刊列表、ACM TODS、ACM TOIS、ACM TKDD、IEEE TKDE）论文数不少于 3 篇；在权威国际会议上报告论文数不少于 10 篇；重要中文期刊（如国家自然科学基金委员会认定）论文数不少于 30 篇；专著出版 2 部；发明专利获得 5 个；争取获得国家/省部级奖励 1 项；争取获得国家优秀青年基金/新世纪人才奖励 1 个；取得应用采纳/示范计划证明或批文 2 个；建设实验计算与演示平台 1 个；博士生培养 20 名。

本课题的研究，不仅可以大大丰富国际上大数据研究的理论体系，提升我国学者在国际大数据研究领域的学术地位，而且将指导我国企业大数据商务与管理实践，进而提升我国企业的信息化水平，有助于实现以信息化带动工业化的国家发展战略。

（二）研究基础和工作条件

1、工作基础（与本课题相关的研究工作积累和已取得的初步研究成果）

申请人及项目组成员多年来一直从事电子商务、数据挖掘、商务智能、信息管理与信息系统、智能决策支持系统等方面的研究与开发工作。近 5 年来，承担与本项目相关的国家重点基础研究发展计划（973 计划）子课题、国家“核高基”科技重大专项子课题各 1 项、国家自然科学基金课题 9 项、国家“863”项目 4 项、企业委托课题多项。在《Marketing Science》、《IEEE Transactions on Software Engineering》、《Journal of Marketing》、《MIS Quarterly》、《Information System Research》、《International Journal of Production Economics》、《IEEE Transactions on Systems, Man, and Cybernetics》、《Decision Support Systems》、《Renewable & Sustainable Energy Reviews》、《Computer & Operations Research》、《European Journal of Operational Research》、《管理科学学报》、《系统工程理论与实践》、《系统工程学报》、《计算机研究与发展》等国内外学术期刊上发表论文 300 余篇，其中 SCI 或 SSCI 收录 80 余篇，EI 收录 120 余篇。授权相关发明专利 3 项，申请相关发明专利（已进入实质审查阶段）7 项。在科学出版社出版学术专著 5 部；获国家科技进步奖二等奖 2 项、教育部自然科学奖一等奖 1 项、安徽省科技进步奖一等奖 3 项、安徽省科技进步奖二等奖 3 项、其他学术奖



励多项。相关研究工作及取得的成果为本课题的研究奠定了必要的理论、方法与技术基础。

主要工作说明如下：

(1) 在大数据分析 with 计算方法及其关键技术方面。项目组主要成员通过承担的国家自然科学基金项目“商务智能中的动态数据挖掘与分形技术研究”、“云计算环境下基于分形数据挖掘技术的商务智能系统的研究”、“不确定信息环境下基于大规模数据的趋势预测和智能决策方法研究”以及国家“863”重点项目“智能配用电信息及通信支持技术与开发”子课题“智能配用电数据预处理技术研究”、“制造业商务智能技术及产品研发”子课题“制造业商务智能支持方法研究”、国家“863”项目“面向制造业售后服务的商务智能关键技术的研究”等研究工作，在面向大规模数据的预处理方法、查询方法、聚类方法、案例推理方法、关联分类方法、蚁群优化方法、面向制造业流数据的分形挖掘方法以及知识融合方法等方面取得了一系列的理论方法上的成果。

1) 论文

- [1] Yezheng Liu, Yuanchun Jiang, Xiao Liu, Shanlin Yang, CSMC: A combination strategy for multi-class classification based on multiple association rules, Knowledge-Based Systems, 21(8), pp 786-793, 2008/12
- [2] Kaile Zhou, Shanlin Yang, Chao Shen, A review of electric load classification in smart grid environment, Renewable & Sustainable Energy Reviews, 24, 103-110, 2013.
- [3] Zhiwei Ni, Junfeng Guo, Li Wang, and Yazhuo Gao, An Efficient Method for Improving Query Efficiency in Data Warehouse, Journal of software, 6(5), 857-865, 2011.
- [4] Chao Fu, Shanlin Yang, An evidential reasoning based consensus model for multiple attribute group decision analysis problems with interval-valued group consensus requirements, European Journal of Operational Research, 223(1), 167-176, 2012.
- [5] Chao Fu, Shanlin Yang, An attribute weight based feedback model for multiple attributive group decision analysis problems with group consensus requirements in evidential reasoning context, European Journal of Operational Research, 212(1), 179-189, 2011.
- [6] Chao Fu, Shanlin Yang, The group consensus based evidential reasoning approach for multiple attributive group decision analysis, European Journal of Operational Research, 206 (3), 601-608, 2010.
- [7] Jiadong Zhang, Jin Xu, Stephen Shaoyi Liao, Aggregating and sampling methods for processing GPS data streams for traffic state estimation[J], IEEE Transactions on Intelligent Transportation Systems, 14(4), 1629-1641, 2013.
- [8] Gang Wang, Jian Ma, Lihua Huang, Kaiquan Xu, Two Credit Scoring Models Based on Dual Strategy Ensemble Trees, Knowledge Based Systems, 26(1), 61-68, 2012.
- [9] Shanlin Yang, Yongwu Zhou, Two-echelon supply chain models: considering duopolistic retailers' different competitive behaviors, International Journal of Production Economics, 103(1), 104-116, 2006.
- [10] Jiadong Zhang, Jin Xu, Stephen Shaoyi Liao, Sampling methods for summarizing unordered vehicle-to-vehicle data streams[J], Transportation Research Part C: Emerging Technologies, 23, 56-67, 2012.
- [11] 杨善林, 李永森, 胡笑旋, 潘若愚, K-Means 算法中的 K 值优化问题研究, 系统工程理论与实践, 26(2), 97-101, 2006.



- [12] 高雅卓, 倪志伟, 倪丽萍, 连续属性上的 OLAP 查询建模方法研究, 情报学报, 30(30), 372-379, 2011.

2) 专著

- [1] 倪志伟, 倪丽萍, 刘慧婷, 贾瑞玉, 动态数据挖掘, 科学出版社, 32 万字, 2010.

3) 专利

- [1] 倪志伟等, 基于数据流的车辆故障诊断系统, 授权号 ZL 2010 0183798.X
[2] 杨善林, 丁静, 罗贺等, 云计算环境下的分类规则挖掘方法, 授权号 2012 1 0203816.5
[3] 杨善林, 凌骥生, 褚伟等, 金属液综合性能在线检测智能检测系统, 授权号 ZL2006 1 0002850.0

4) 奖励

- [1] 面向企业的智能决策理论与方法.教育部自然科学一等奖.杨善林,周永务,胡小建,刘心报,梁昌勇,刘业政,朱卫东,任明仑,马溪骏.2008.
[2] 铸铁材质参数液态在线智能检测与质量控制系统. 国家科技进步奖二等奖. 杨善林, 凌骥生, 褚伟, 马溪骏, 周永务, 左闯, 任明仑, 刘心报, 余本功, 潘若愚. 2006.
[3] 基于知识发现与定性推理的智能决策支持系统. 安徽省自然科学二等奖. 杨善林, 周永务, 刘心报, 黄梯云, 朱卫东, 梁昌勇, 任明仑, 马溪骏. 2004.
[4] 付超博士论文《基于置信信念函数的群决策过程研究》2011 年被评为安徽省优秀博士学位论文。

(2) 在大数据获取系统、支撑平台与可视化技术方面。项目组主要成员通过承担的国家自然科学基金项目“云计算环境下的多源信息服务系统研究”、国家“863”重点项目“集团企业云制造服务平台共性关键技术研究”, 以及企业委托课题“警务数据智慧应用平台(一期)”、“基于大数据的个性化交通信息推荐系统”等研究工作, 在面向大规模数据的获取方法、管理技术、可用性评估、可视化技术、基于云计算的信息系统体系结构和平台构建策略等方面取得了一系列的理论方法和关键技术上的成果。

1) 论文

- [1] Shuai Ding, Xijun Ma, Shanlin Yang, A software trustworthiness evaluation model using objective weight based evidential reasoning approach, Knowledge and Information Systems, 33(1), 171-189, 2012.
[2] Shuai Ding, Shanlin Yang, Youtao Zhang, et al, Combining QoS prediction and customer satisfaction estimation to solve cloud service trustworthiness evaluation problems[J], Knowledge-Based Systems, 56, 216-225, 2014.
[3] Xiao Liu, Yun Yang, Yuanchun Jiang, Preventing Temporal Violations in Scientific Workflows: Where and How, IEEE Transactions on Software Engineering, 37(6), 805-825, 2011.(2)
[4] Shuai Ding, Chenyi Xia, Kaile Zhou, Shanlin Yang, Jennifer S. Shang, Decision Support for Personalized Cloud Service Selection through Multi-Attribute Trustworthiness Evaluation[J], PloS one, 9(6), e97762, 2014.
[5] 杨善林, 丁帅, 褚伟, 一种基于效用和证据理论的可信软件评估方法, 计算机研究与发



展, 48(7), 1152-1159, 2009.

- [6] 杨善林, 罗贺, 丁帅, 基于云计算的多源信息服务系统研究综述, 管理科学学报, 15(5), 83-96, 2012.
- [7] 杨善林, 刘业政, 李亚飞, 基于 Rough Sets 理论的证据获取与合成方法[J], 管理科学学报, 8(5), 69-75, 2006.
- [8] 杨善林, 丁帅, 付超, 考虑信息源相关的软件可信性评估模型, 中国管理科学, 17(6), 163-169, 2009.
- [9] 丁帅, 鲁付俊, 杨善林, 夏成遗, 一种需求驱动的软件可信性评估及演化模型, 计算机研究与发展, 48(4), 647-655, 2011.

2) 专利

- [1] 杨善林, 凌骥生, 褚伟等, 金属液综合性能在线检测用传感器, 授权号 ZL2006 1 0002851.5

3) 奖励

- [1] 医院信息系统. 安徽省科技进步二等奖. 杨善林, 余本功, 刘业政, 任明仑, 马溪骏. 2003.

(3) 在基于网络大数据的商业情报分析方面。项目组主要成员通过承担的国家重点基础研究发展计划(973 计划)项目子课题“社交网络分析与网络信息传播的基础研究: 社交网络群体行为与群体情感理论研究”、教育部人文社科项目“面向互联网大数据的舆论领袖话语影响力评估和治理方法研究”、“网络社区舆情及舆论领袖识别方法研究”、“在线社会性网络的信息传播及其对客户关系管理的影响”等研究工作, 在面向大规模数据的短文本分析方法、热点话题、观点与情感演化等方面取得了一系列的理论方法上的成果。

1) 论文

- [1] Gang Wang, Jianshan Sun, Jian Ma, Kaiquan Xu, Jibao Gu, Sentiment Classification: The Contribution of Ensemble Learning, Decision Support Systems, 57, 77-93, 2014.
- [2] Liyang Tang, Zhiwei Ni, Hui Xiong, Hengshu Zhu, Locating targets through mention in Twitter[J], World Wide Web, 1-31, 2014.
- [3] Raymond Y.K. Lau, Chunping Li, Stephen Shaoyi Liao, Social Analytics: Learning Fuzzy Product Ontologies for Aspect-Oriented Sentiment Analysis, Decision Support Systems, 2014
- [4] Kaiquan Xu, Stephen Shaoyi Liao, Raymond Lau and Leon Zhao, Effective Active Learning Strategies for the Use of Large-Margin Classifiers in Semantic Annotation: An Optimal Parameter Discovery Perspective[J], INFORMS Journal on Computing, 2014.
- [5] Raymond Y.K. Lau., Stephen Shaoyi Liao, et al., Text Mining and Probabilistic Language Modelling for Online Review Spam Detection, ACM Transactions on Management Information Systems, 2011, 2(4): 1-30.
- [6] Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li and Yuxia Song. Mining comparative opinions from customer reviews for competitive intelligence. Decision Support Systems, 2011, 50(4): 743-754
- [7] Dongyuan Lu, Qiudan Li, Stephen Shaoyi Liao, A Graph-based Action Network Framework to Identify Prestigious Members through Member's Prestige Evolution, Decision Support Systems, 2012, 53(1): 44-54.
- [8] Jianmin He, Mengna Hu, Mingguang Shi, Yezheng Liu, Research on the Measure Method



of Complaint Theme Influence on Online Social Network, Expert Systems with Applications, 41(13), 6039-6046, 2014.

- [9] 刘业政, 周云龙, 无尺度网络平均路径长度的估计, 系统工程理论与实践, 34(6), 1566-1571, 2014.
- [10] 孙春华, 刘业政, 基于产品特征词关系识别的评论倾向性合成方法, 情报学报, 32(8), 844-852, 2013.
- [11] 孙春华, 刘业政, 彭学仕, 一种含强度的基准词选择和词汇倾向性判别方法, 情报学报, 30(12), 1261-1267, 2011.
- [12] 刘业政, 杜亚楠, 姜元春, 杜非, 基于热度曲线分类建模的微博热门话题预测, 模式识别与人工智能. (已录用)
- [13] 何建民, 常传武, 刘业政, 客户网上参与产品开发的“动机-行为”模型研究, 中国管理科学, 19(5), 173-181, 2011.

2) 专著

- [1] 刘业政, 姜元春, 张结魁, 网络消费者行为: 理论方法及应用, 科学出版社, 30 万字, 2011.

(4) 在大数据驱动的个性化信息服务方面。项目组主要成员通过承担的国家自然科学基金项目“基于模体挖掘面向在线社交网络中虚拟社区的群推荐系统研究”、“基于主动协作关系的网络用户建模及模型优化方法研究”、“基于网络消费者信息需求模型的网站导航问题研究”、“具备情境感知特性的项目知识推荐系统方法研究”、“软件服务化背景下的云服务推荐方法研究”等研究工作, 在基于多源大数据的网络用户建模、个性化推荐算法、个性化营销策略等方面取得了一系列的理论方法上的成果。

1) 论文

- [1] Yuanchun Jiang, Jennifer Shang, Chris F. Kemerer, Yezheng Liu, Optimizing E-tailer Profits and Customer Savings: An Online Dynamic Bundle Pricing Model, Marketing Science, 30(4), 737-752, 2011.
- [2] Yuanchun Jiang, Jennifer Shang, Yezheng Liu, Maximizing Customer Satisfaction through an Online Recommendation System: A Novel Associative Classification Model, Decision Support Systems, 48(3), 470-479, 2010.
- [3] Yuanchun Jiang, Jennifer Shang, and Yezheng Liu, Optimizing shipping-fee schedules to maximize e-tailer profits, International Journal of Production Economics, 146(2), 634-645, 2013.
- [4] Yuanchun Jiang, Yezheng Liu, Optimization of Online Promotion: A Profit-Maximizing Model Integrating Price Discount and Product Recommendation, International Journal of Information Technology & Decision Making, 11(5), 1-22, 2012.
- [5] 刘业政, 叶俊杰, 蒋玮, 徐俊芬, 基于图上随机游动的 Web 页面布局可用性评价, 管理科学学报, 16(9), 75-81, 2013.
- [6] Raymond Lau, Stephen Shaoyi Liao, K.F. Wong, and Dickson Chiu, Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions, MIS Quarterly, 2012, 36(4): 1239-1268.
- [7] Heng Tang, Stephen Shaoyi Liao, Sherry Sun, A Prediction Framework Based on Contextual Data to Support Mobile Personalized Marketing, Decision Support Systems, 2013, 56: 234-246.



2) 奖励

- [1] 姜元春博士学位论文《基于智能优化方法的个性化营销策略研究》获安徽省第四届优秀博士论文奖, 2013 年度全国优秀博士论文提名奖。

3) 专利

- [1] 刘业政, 姜元春, 王锦坤等, 一种基于推荐概率融合的混合推荐方法, 申请号 201310637512.4.

(5) 在与大数据相关的实际应用方面。项目组主要成员完成了国家“核高基”科技重大专项“大型网络应用及服务平台研制与示范”子课题“大型网络应用及服务平台推荐方法与测试策略研究”、企业委托项目“警务数据智慧应用平台”、“基于大数据的个性化交通信息推荐系统”、“中国联通合肥分公司离网客户数据挖掘及客户发展定位研究”、“江苏电网配用电数据分析”、“奇瑞汽车售后服务数据分析”、“基于网络评论数据分析的汽车品牌管理系统”等研究工作, 将取得的相关理论成果与实践应用相结合, 为组织管理决策提供支持。从应用实践的研究过程中也进一步加深了对大数据环境下商务管理问题的理解。

1) 奖励

- [1] 面向供应链的制造工程协同优化与集成制造执行系统, 2012 年安徽省科学技术奖一等奖, 获奖人员依次为: 杨善林, 余本功, 胡小建, 鲁付俊, 高立新, 任明仑, 马英, 杨颖, 刘刚, 吴卓亚.
- [2] 轿车整车自主开发系统的关键技术研究及其工程应用. 国家科技进步二等奖. 尹同耀, 杨善林, 陆建辉, 鲁付俊, 高立新, 梁昌勇, 刘心报, 余本功, 沈浩杰, 王磊. 2008.
- [3] 离散制造业生产优化与执行系统. 安徽省科技进步一等奖. 杨善林, 刘心报, 李小宇, 刘林, 安涛, 许安平, 程浩, 汤金春, 裴凤, 毛征. 2008.
- [4] 轿车整车开发系统平台、流程优化与工程管理系统. 中国机械工业科学技术奖一等奖. 尹同耀, 杨善林, 陆建辉, 鲁付俊, 高立新, 梁昌勇, 刘心报, 余本功, 沈浩杰, 金德全, 王磊, 胡小建, 杨颖, 罗贺, 程飞. 2007.
- [5] 李军, “不确定情景下路径选择理论与应用研究”, 四川省科技进步二等奖, 2008
- [6] 李军, “不确定情景下路径选择理论与应用研究”, 成都市科技进步三等奖, 2008
- [7] 李军, “城市货运卡车集中优化调度研究”, 四川省科技部, 四川省科技进步二等奖, 1995

2) 鉴定成果

“警务数据智慧应用平台”结合犯罪者行为特征, 基于警务系统数据、社会系统数据以及社交网络数据, 借助社交网络分析方法和大数据分析方法, 一方面分析犯罪嫌疑人涉及的人、地、物、组织、案(事)等关联关系, 判断犯罪嫌疑人的犯罪可能性, 为案件侦破提供决策支持; 另一方面, 分析人群的行为轨迹(包括网络交往、言论; 现实生活中的交往、吃、住、行等), 发现潜在的犯罪行为。



鉴 定 意 见
1、提供的鉴定资料齐全、完整，符合鉴定要求。
2、该项目针对警务数据多源性、分散式、碎片化等特点，从警务实际应用需求出发，综合利用人、案（事）件、物品、地点、组织、时间等要素，研发出一套多并发警用 PETL 工具，建立了整合公安内部业务数据资源、社会基础数据资源和互联网开放数据资源的多维数据中心，为警务数据的智能分析奠定了基础。
3、该项目针对公安数据中各要素间复杂的关联关系，提出了以人为核心的主题分析策略，建立了人员轨迹、关系、涉物、涉案（事）等关联分析模型，可分析出与该人相关的信息，并可视化展示分析结果。
4、该项目构建了基于图形展现技术的数据处理与分析方法库。基于公安基础数据和关联分析模型，提出了以案发地标为中心的同类案件嫌疑人线索排查方法。
5、该平台软件通过了安徽省软件评测中心的测试，经实际应用，提高了办案效率，取得了较好的社会和经济效益。
该项目总体技术达到国内先进水平，在基于大数据的情报分析方面处于国内领先水平，鉴定委员会一致同意通过该项目成果鉴定。
建议：进一步优化数据挖掘技术，加大产品推广力度。
鉴定委员会主任：_____ 副主任：_____
2014年 4 月 29 日

2、工作条件（已具备的实验条件，尚缺少的实验条件和拟解决的途径，包括利用重点研究基地的情况，如国家实验室、国家重点实验室和大科学工程等）

合肥工业大学管理学院拥有“管理科学与工程”一级学科国家重点学科和“工商管理”一级学科博士学位授权点、博士后流动站；建有“智能决策与信息系统技术”国家地方联合工程研究中心、“过程优化与智能决策”教育部重点实验室、“智能决策与信息系统技术”教育部工程研究中心、“知识经济与企业管理创新”安徽省人文社会科学重点研究基地、“决策科学与信息系统技术”教育部创新团队。同时，还建有电子商务实验室、商业智能实验室、决策科学与技术实验室等，拥有价值 1000 多万元的仪器设备、可供本项目使用的软件以及先进的网络系统。

项目团队与美国匹兹堡大学、佛罗里达州立大学、英国曼彻斯特大学、布鲁内尔大学、澳大利亚斯韦本科技大学、加拿大多伦多大学、麦克马斯特大学以及香港城市大学等建立了经常性的学术交流机制。

合肥工业大学有良好的校园环境和先进的互联网络，有丰富的图书资料，有浓厚的学术氛围，还有支持自然科学基金项目研究工作的政策规定。



项目组与多家电子商务企业和制造企业有长期科研合作关系，与芜湖奇瑞、江淮汽车、百度公司、阿里巴巴、安徽联通、安徽电信等企业建立了良好的合作关系，为开展本项目的实证研究提供了环境和平台。

3、申请人和项目组主要参与者正在承担的国家自然科学基金和国家其他科技计划项目情况（需注明计划名称、项目名称和编号、起止年月、与本课题的关系及负责的内容等）。

3.1 在研项目名称

（1）国家自然科学基金，基于模体挖掘面向在线社交网络中虚拟社区的群推荐系统研究（71371062），2014.01-2017.12，刘业政主持，负责总体规划、多模社交网络建模。

（2）国家重点基础研究发展计划（973 计划），社交网络分析与网络信息传播的基础研究：社交网络群体行为与群体情感理论研究（2013CB329603），2013.01-2017.12，主要参与者：杨善林，刘业政；杨善林负责合肥工业大学承担的研究内容的总体规划，刘业政负责合肥工业大学承担的研究内容的具体执行、基于短文本的社交网络用户行为分析方法。

（3）国家自然科学基金重点项目，云计算环境下的多源信息服务系统研究（71131002），2012.1-2016.12，杨善林主持，负责总体规划与应用示范设计。

（4）国家自然科学基金，云计算环境下基于分形数据挖掘技术的商务智能系统的研究（71271071），2013.01 -2016.12，倪志伟主持，负责项目总体规划、商务智能系统的体系结构、云计算环境下的数据管理技术。

（5）国家自然科学基金，具有竞争行为的公路合作运输决策研究（71271178），2013.01-2016.12，李军主持，负责项目总体规划、竞争模型构建。

（6）教育部人文社科规划基金项目，面向互联网大数据的舆论领袖话语影响力评估和治理方法研究（14YJA630015），2014.01-2016.12，何建民主持，负责影响力度量方法。

（7）国家自然科学基金青年基金，在线替代品系列销售的动态协同价格策略研究（71302064），2014.01-2016.12，姜元春主持，负责智能优化算法。

（8）国家自然科学基金青年基金，软件服务化背景下的云服务推荐方法研究（71201042），2013.01-2015.12，丁帅主持，负责云服务推荐算法研究。

（9）国家自然科学基金青年基金，不确定信息环境下基于大规模数据的趋势预测和智能决策方法研究（71101041），2012.1-2014.12，王晓佳主持，负责不确定性建模和数据趋势预测方法。

3.2 与本项目的关系

上述项目针对互联网等典型大数据环境下的不确定性信息处理、系统建模、知识发现、信息融合、商务智能、个性化推荐、资源的优化调度等问题进行了系统研究，对大数据的获取方法、管理技术和分析方法、大规模不确定性数据的趋势预测方法、云计算环境下基



于分形技术的系统建模方法和数据挖掘方法、云计算环境下的资源分配方法有了较为系统的理解和把握。这些内容是开展大数据环境下数据可用性评估、大数据系统建模、大数据分析方法和大数据支撑平台建设以及大数据驱动的商业情报分析、个性化信息推荐等研究工作的基础，本课题也能够进一步促进上述项目的深入开展。

4、完成国家自然科学基金项目情况：(对申请人负责的前一个已结题项目（项目名称及批准号）完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该已结题项目研究工作总结摘要（限500字）和相关成果的详细目录）。

4.1 项目名称与编号

申请人此前已完成两项国家自然科学基金项目，其中2009年12月结题的项目“基于网络消费者信息需求模型的网站导航问题研究”（70672097，2007.01-2009.12）在结题项目绩效评估中被评为“优”。

最近结题的项目完成于2013年12月，“基于主动协作关系的网络用户建模及模型优化方法研究（71071047，2011.01-2013.12）”。下面对本项目的完成情况、后续研究进展及与本课题的关系加以详细说明。

4.2 项目总体完成情况

该项目主要研究了（1）在线社会性网络服务中的用户关联关系；（2）在线评论与消费者行为偏好间的关系；（3）集成新兴信息服务的网络用户建模；（4）考虑认知风格的个性化捆绑推荐策略。经过三年的研究，取得了预期效果，在《Marketing Science》国际顶级期刊以及《International Journal of Production Economics》、《Journal of Information Technology & Decision Making》、《Journal of Multi - Criteria Decision Analysis》、《管理科学学报》、《系统工程理论与实践》、《系统工程学报》、《管理工程学报》、《中国管理科学》、《情报学报》等其他国内外著名期刊上共发表论文30篇，其中被SCI收录2篇，SSCI收录1篇，EI收录6篇，基金委指定管理科学部A类期刊10篇，26篇中文期刊论文均为CSSCI或CSCD收录期刊；在《4th International Conference on Intelligent Decision Technologies (IDT)》等学术会议上共发表论文3篇，被EI收录1篇，ISTP收录1篇；出版专著1部，获安徽省教学成果特等奖1项，出版“十一五”国家规划教材《电子商务概论》1部，并入选国家“十二五”规划教材；承办了1次全国学术会议，支持项目组成员参加了16人次境内、2人次境外的国际、国内学术会议，邀请了2名国外学者来华开展合作研究；支持了8名博士生和10名硕士生的学位论文，其中4人已取得博士学位，10人已取得硕士学位，1篇博士论文获全国优秀博士论文提名奖和安徽省优秀博士论文，2篇硕士论文获安徽省优秀硕士论文。



4.3 研究工作总结摘要

本项目将体现消费者主动协作关系的在线社会性网络服务、在线评论等新兴信息源引入用户建模,利用消费者行为学、认知心理学、社会性网络分析、行为运筹学等理论方法,围绕在线社会性网络服务中的用户关联关系、在线评论与消费者行为偏好间的关系、集成新兴信息服务的网络用户建模、考虑认知风格的个性化捆绑推荐策略等开展了系统的研究。项目组通过规范的实证研究方法揭示了在线社会性网络服务的影响因素及在线社会性网络对用户行为的影响;针对大规模复杂网络的平均路径长度计算中存在的问题,提出了一种树形结构模型,给出了无尺度网络、随机网络和小世界网络的平均路径长度的近似计算公式,并分析了网络规模和节点间连接方式对平均路径长度的影响。构建了在线评论有用性影响因素模型,揭示了不同消费者对同一产品产生不同的使用感受并发表不同的在线评论的原因;针对在线评论的语义描述方法、词汇倾向性判别方法和评论倾向性合成方法等关键问题,提出了含强度的基准词选择和词汇倾向性判别方法和基于产品特征词关系识别的评论倾向性合成方法,为较好的解决用户建模过程如何集成在线评论信息提供了解决方案。综合利用多信息源,提出了基于SVM的不均衡数据分类方法、两阶段参数自适应蚁群算法、基于特征聚类的相似领域上下文学习模型、面向领域的数据包络分析方法等用户建模方法。针对页面内容展示对购买行为的影响,基于消费者认知风格,构建了商品展示页面、内容和布局可用性评价模型,揭示了网络消费者认知风格对页面复杂度、信息关注度和布局可用性偏好的影响;构建了在用户偏好模型、用户认知风格、产品整体满意度、企业内部信息(库存、成本、利润率)等条件的独立或联合约束下产品捆绑、产品展示、个性化定价等个性化捆绑推荐的最优化模型,构建了基于用户偏好、购买行为和配送方式选择行为的配送规划模型,并设计了基于遗传算法等智能优化方法求解相应的模型。本项目的研究对发展新的网络用户建模理论,提高用户满意度和企业经营效益具有重要意义。

4.4 项目研究成果目录

[1]Jiang, Yuanchun、Shang, Jennifer、Kemerer, Chris F.、Liu, Yezheng, Optimizing E-tailer Profits and Customer Savings: Pricing Multistage Customized Online Bundles, Marketing Science, 30(4), pp 737-752, 2011/8JUL-AUG

[2]Jiang, Yuanchun、Liu, Yezheng, Optimization of Online Promotion: A Profit-Maximizing Model Integrating Price Discount and Product Recommendation, Journal of Information Technology & Decision Making, 11(3), pp 961-982, 2012/10/1

[3]Yuanchun Jiang、Jennifer Shang、Yezheng Liu, Optimizing shipping-fee schedules to maximize e-tailer profits, International Journal of Production Economics, 146(2), pp 634-645, 2013

[4]刘业政、叶俊杰、蒋玮、徐俊芬, 基于图上随机游动的 Web 页面布局可用性评价, 管理科学学报, 09 期, pp 75-81, 2013/9/15



- [5]刘业政、周云龙, 无尺度网络平均路径长度的估计, 系统工程理论与实践, 2014.6, 1566-1571
- [6]孙春华、刘业政, 基于产品特征词关系识别的评论倾向性合成方法, 情报学报, 32(8), pp 844-852, 2013
- [7]Chen Sifeng、Jiang Yuanchun、Liu Yezheng, Cost constrained mediation model for analytic hierarchy process negotiated decision making , Journal of Multi - Criteria Decision Analysis, 19(1-2), pp 3-13, 2012
- [8]张启平、刘业政、刁翠霞, 面向社会经济发展水平的钢铁生产效率 DEA 动态评价 , 系统工程理论与实践, 11 期, pp 2577-2584, 2012
- [9]张启平、刘业政、李勇军, 考虑受益性的固定成本分摊 DEA 纳什讨价还价模型 , 系统工程理论与实践, 2014.3, 756-768
- [10]丁正平、刘业政, 存在搭便车时双渠道供应链的收益共享契约 , 系统工程学报, 03 期, pp 370-376, 2013/6/15
- [11]刁翠霞、陈思凤、刘业政, 基于 SVM 求解不均衡数据集分类的主观权重约束方法 , 管理工程学报, 03 期, pp 146-150, 2012
- [12]何建民、常传武、刘业政, 客户网上参与产品开发的“动机-行为”模型研究, 中国管理科学, 05 期, pp 173-181, 2011
- [13]刘业政、张启平、姜元春, 决策单元交叉效率的自适应群评价方法, 中国管理科学 (录用)
- [14]孙春华、刘业政、彭学仕, 一种含强度的基准词选择和词汇倾向性判别方法, 情报学报, 30(12), pp 1261-1267, 2011
- [15]陈思凤、刘业政 , 混合状态变量和过程变量的风险评估模型, 控制与决策, 02 期, pp 309-312, 2013/2/15
- [16]凌海峰、王西山 , 求解柔性作业车间调度问题的两阶段参数自适应蚁群算法, 中国机械工程, 24(24), pp 3380-3385, 2013
- [17]代宝、刘业政, 基于技术接受模型和感知流行性的 SNS 使用意愿研究, 科技进步与对策, 24 卷, pp 1-6, 2012/12/1
- [18]王平、代宝, 消费者在线评论有用性影响因素实证研究, 统计与决策, 2012(2), pp 118-120, 2012/2/1
- [19]叶俊杰、刘业政、蒋玮, Web 环境下认知风格对商品信息关注度的影响研究, 图书情报工作, 06 期, pp 95-101, 2012
- [20]何建民、孙庆庆、刘业政 , 复合参数异步检测网络热点话题方法研究, 情报杂志, 10 期, pp 145-150, 2011
- [21]蒋玮、叶俊杰、刘业政 , 消费者认知风格对 Web 页面复杂度偏好影响的实证研究, 情报杂志, 07 期, pp 178-184, 2011



- [22]魏婧、刘业政、邢小云，在线社会化网络的口碑传播研究，情报杂志, 04 期, pp 34-38+54, 2011
- [23]邢小云、魏婧，在线社会网络结构演化及其对口碑传播的影响研究，现代图书情报技术, 27(9), pp 60-65, 2011
- [24]凌海峰、余笪，基于聚类的 Web 用户会话识别优化方法，计算机应用研究, 29(8), pp 2862-2864, 2012
- [25]宋仙磊、刘业政、陈思凤，基于周期项方法选择的季节性时序预测，计算机工程, 21 期, pp 131-132+135, 2011
- [26]张可、凌海峰，基于均匀设计和混沌理论的蚁群算法参数调整，计算机工程, 38(14), pp 141-143, 2012
- [27]徐俊芬、叶俊杰、刘业政，基于相似领域共享特征的分类学习模型，计算机工程与应用
- [28]凌海峰、曹荣涛，基于混合蚁群算法的 Web 用户会话聚类，计算机工程与应用
- [29]凌海峰、王浩，细菌觅食算法与 K-means 结合的 Web 用户会话聚类，计算机工程与应用, 48(36), pp 121-124, 2012
- [30]何建民、刘明星，面向特定主题的客户抱怨文本分类识别方法，合肥工业大学学报(自然科学版), 33(12), pp1884-1888, 2010
- [31]Liu Yezheng、Ding Zhengping，Revenue Sharing Contract in Dual Channel Supply Chain in Case of Free Riding，4th International Conference on Intelligent Decision Technologies (IDT), 2012/5/22-2012/5/25, pp 459-469, Gifu, 2012，EI,ISTP
- [32]姬建睿、姜元春、刘业政，中文评论中产品特征与情感识别研究，信息系统协会中国分会第四届学术年会，2011, 2011/1/1-2011/1/2, 2011/1/1
- [33]王锦坤、姜元春、刘业政，面向实体超市的商品推荐系统架构，信息系统协会中国分会第五届学术年会(CNAIS 2013), 2013/10/17-2013/10/19, pp 431-434, 贵阳, 2013/10/18
- [34]杨善林、刘业政，以创新人才培养为目标的教学团队建设，安徽省教育厅，安徽省教学成果特等奖，省级, 2012/12
- [35]刘业政，电子商务概论(第 2 版)，高等教育出版社, 2012/8/1
- [36]刘业政、姜元春、王锦坤，一种基于概率推荐融合的混合推荐方法，中国，中国国家专利局, 201310637512.4, 专利

4.5 后续研究进展

通过该项目的研究,项目组发现随着 Web2.0 应用以及其他各种类型的社会媒体的发展,在线社交网络已经成为人们网络生活的最主要平台。为此,在已有研究基础上,秉承已有的研究思路,以在线社交网络平台为核心,从多个方面继续开展研究:(1)面向虚拟社区的群推荐系统研究。与面向消费者个体的推荐系统相比,由于社会影响对群成员的最终选



择产生重要作用,使得群推荐系统提供的信息更具价值,具有更高的营销效率。主要研究内容:基于社会归属网络的社交网络表示模型,以及模型的结构演化和功能特征;团渗透法扩展策略以及基于模体渗透法的具有共同兴趣的虚拟社区发现方法;虚拟社区形成过程中选择效应和社会影响效应的不同影响;个体偏好的获取策略、群偏好集结方法与模型优化策略,以及群推荐候选方案的产生方法、群推荐结果的协调方法以及系列群推荐的最优策略。(2) 社交网络群体行为与群体情感理论研究。主要研究内容:社交网络用户的个体人格结构、心理因素和经历背景等要素与其行为特征、情感立场的关系;面向社交网络中海量碎片化的个体行为信息的多属性、多维度的统计分析方法和机器学习方法;网络群体的共性认知空间和多维属性的群体情感建模方法;具有共性情感认知空间的群体间的互动对群体情感的影响以及群体情感演化的计算方法。(3) 基于群营销的品牌社会化传播与推广研究。主要研究内容:社交网络环境下的品牌传播与扩散机理;品牌社会化传播与推广优化方法;基于消费者品牌认知度的品牌关系评价方法;品牌关注度分析及其演化规律、顾客的品牌情感依赖度分析及其演化规律、品牌危机管理及其对品牌资产的影响等。

4.6 与本项目的关系

① 通过该项目的研究,项目组对网络环境下的基于网络爬虫的信息收集方法以及数据预处理方法、多源数据融合方法、面向在线评论的短文本数据处理、基于 SVM 和基于关联分类的数据挖掘方法、基于蚁群优化的智能优化方法有了较系统的理解和把握,这些内容为本课题的研究提供了基础。

② 通过该项目的研究,项目组对网络消费者的多维度兴趣偏好以及相应的建模方法、个性化推荐方法有了较系统的理解和把握,对完成大数据驱动的网络用户建模及个性化信息推荐方法提供了重要的基础。

③ 后续研究进展主要围绕在线社会性网络等典型大数据环境下的信息传播、信息融合、文本主题分析与情感计算等方面展开,并积累了多种社交网络用户分析方法,初步构建了基于大数据的汽车品牌管理系统平台,也为本课题的研究形成了重要支撑。

④ 通过该项目的研究工作,形成了一支年龄和知识结构合理的研究队伍,为本项目的研究打下了较坚实的人力资源基础;同时,我们还和奇瑞汽车、江淮汽车、百度、阿里巴巴、太平洋汽车网等企业建立了良好的关系,并通过不同渠道收集了大量的与本项目相关的研究数据:在线社会性网络的成员关系数据(7年)、新浪微博数据、太平洋汽车网在线评论数据、美国家庭网站访问与在线交易数据(5年)等。



（三）申请人和项目组主要参与者简介（在读研究生除外）。

1、刘业政，合肥工业大学，管理学院，博士，教授，博士生导师。本课题负责人，并具体负责“专题六：案例研究：基于大数据的汽车品牌管理系统”研究。

1.1 个人简介

刘业政教授是“新世纪百千万人才工程”国家级人选，国务院政府津贴专家。主要从事电子商务、商务智能、智能决策支持系统等领域的教学与研究工作。先后主持国家自然科学基金3项、国家“863”重点项目子课题1项、国家“核高基”科技重大专项子课题1项以及教育部博士点科研基金、安徽省自然科学基金等省部级科研项目5项，参加国家“973”计划项目、自然科学基金重点项目等国家级、省部级课题10余项，同时还主持或参加了多项横向课题的开发与研究工作。先后获省部级科技进步一等奖2项、二等奖1项、三等奖1项，获安徽省第四届青年科技创新奖；获安徽省教学成果特等奖3项、国家教学成果二等奖3项；出版或参与出版专著3部，主编“十二五”国家规划教材1部，主编或副主编“十一五”国家规划教材2部，发表学术论文100余篇。现为合肥工业大学电子商务研究所所长，教育部电子商务专业教学指导委员会委员，教育部过程优化与智能决策重点实验室副主任。

1.2 受教育经历（从大学本科开始，按时间倒排序）

1997/09—2002/06，合肥工业大学，管理学院信息管理系，博士

1988/09—1991/06，合肥工业大学，应用物理系，硕士

1982/09—1986/06，合肥工业大学，应用物理系，学士

1.3 工作经历（按时间倒排序）

（1）工作经历

2007/08—至今，合肥工业大学，管理学院电子商务系，教授

2007/03—2007/08，McMaster University, DeGroote 商学院，访问学者

2006/08—2007/03，Saint Mary's University, Sobey 商学院，访问学者

2003/01—2006/08，合肥工业大学，管理学院电子商务系，教授

2000/01—2002/12，合肥工业大学，管理学院信息管理系，副教授

1997/09—2000/01，合肥工业大学，管理学院信息管理系，讲师

（2）主持和参加的研究课题

[1] 基于模体挖掘面向在线社交网络中虚拟社区的群推荐系统研究，国家自然科学基金，2014.01-2017.12，主持

[2] 社交网络分析与网络信息传播的基础研究：社交网络群体行为与群体情感理论研究，



国家重点基础研究发展计划（973 计划），2013.01-2017.12，主要参与者

- [3] 基于主动协作关系的网络用户建模及模型优化方法研究，国家自然科学基金，2011.01-2013.12，主持
- [4] 大型网络应用及服务平台研制与示范—大型网络应用及服务平台个性化推荐方法研究，国家“核高基”科技重大专项子课题，2011.01-2012.12，主持
- [5] 警务数据智慧应用平台（一期），安徽新华博委托课题，2010.01-2014.03，主持
- [6] 新兴信息服务环境下的网络用户建模及模型优化方法研究，高等学校博士点基金，2010.01-2012.12，主持
- [7] 集成 OSN 的网络用户建模方法研究，留学回国人员科研启动基金，2010.01-2012.12，主持
- [8] 淘宝网营销工具有效性分析，淘宝网委托课题，2010.08-2012.12，主持
- [9] 制造业商务智能技术及产品研发—制造业商务智能支持方法研究，国家高技术研究发展计划（863 计划）重点项目子课题，2008.03-2010.09，主持
- [10] 群决策理论与方法研究，国家自然科学基金重点项目，2007.01-2010.10，参加者
- [11] 基于网络消费者信息需求模型的网站导航问题研究，国家自然科学基金，2007.01-2009.12，主持
- [12] 金属液综合性能在线智能检测系统，信息产业部电子发展基金，2004.10-2006.10，参加者
- [13] 基于知识发现的综合决策支持系统研究，国家自然科学基金，2000.01-2002.12，参加者

1.4 近五年发表的与本项目相关的主要论著

（1）期刊论文

- [1] Jianmin He, Mengna Hu, Mingguang Shi, **Yezheng Liu**, Research on the Measure Method of Complaint Theme Influence on Online Social Network, Expert Systems with Applications, 41(13), 6039-6046, 2014.
- [2] Yuanchun Jiang, Jennifer Shang, **Yezheng Liu**, Optimizing Shipping-Fee Schedules to Maximize E-tailer Profits, International Journal of Production Economics, 146(2), 34-645, 2013.
- [3] Haifeng Ling, Hai wang, **Yezheng Liu**, The convergence of ant colony optimization with an adaptive pheromone evaporation rate, ICIC Express Letters, 7(10), 2773-2778, 2013.
- [4] Yuanchun Jiang, **Yezheng Liu**, Optimization of Online Promotion: A Profit-Maximizing Model Integrating Price Discount and Product Recommendation, International Journal of Information Technology & Decision Making, 11(5), 1-22, 2012.
- [5] Sifeng Chen, Yuanchun Jiang, **Yezheng Liu**, Cost Constrained Mediation Model for AHP Negotiated Decision Making, Journal of Multi-Criteria Decision Analysis, 19(1-2), 3-13, 2012.



- [6] Yuanchun Jiang, Jennifer Shang, Chris F. Kemerer, **Yezheng Liu**, Optimizing E-tailer Profits and Customer Savings: Pricing Multistage Customized Online Bundles, *Marketing Science*, 30(4), 737-752, 2011.
- [7] Yuanchun Jiang, **Yezheng Liu**, Xiao Liu, Shanling Yang, Integrating classification capability and reliability in associative classification: A β -stronger model, *Expert Systems with Applications*, 37(5), 3953-3961, 2010.
- [8] Yuanchun Jiang, Jennifer Shang, **Yezheng Liu**, Maximizing customer satisfaction through an online recommendation system: A novel associative classification model, *Decision Support Systems*, 48(3), 470-479, 2010.
- [9] Yong Zhao, **Yezheng Liu**, How to Change an Organization to Fit the Dynamic Environment: A Case Study on a Telecom Company of China[J], *International Journal of Business and Management*, 5(5), 226, 2010.
- [10] **Yezheng Liu**, Jiekui Zhang, Shanlin Yang, Research on Consumers' Cross-site Transfer Behavior Based on Complex Network Theory, *The Innovative Computing, Information and Control Express Letters*, 3(4), 1359-1364, 2009.
- [11] **Yezheng Liu**, Zhengping Ding, Yufei Yuan, Behaviors of Online Financial Institutions Confronted with Phishing: A Game Theoretic Analysis, *Journal of University of Electronic Science and Technology of China*, 38(11), 37-44, 2009.
- [12] **Yezheng Liu**, Yuanchun Jiang, Xiao Liu, Shanlin Yang, CSMC: A combination strategy for multi-class classification based on multiple association rules, *Knowledge-Based Systems*, 21(8):786-793, 2008.
- [13] **Yezheng Liu**, Haifeng Ling, Shanlin Yang, A Hybrid Ant Colony Clustering of Web Browsing Patterns for User Navigation Recommendation, *International Journal of Knowledge and Systems Sciences*, 5(1), 62-66, 2008.
- [14] Haifeng Ling, Shanlin Yang, **Yezheng Liu**, Ant colony approach for navigation pattern mining, *Journal of system science and information*, 5(1), 7-13, 2007.
- [15] 刘业政, 周云龙, 无尺度网络平均路径长度的估计, *系统工程理论与实践*, 34(6), 1566-1571, 2014.
- [16] 张启平, 刘业政, 李勇军, 考虑受益性的固定成本分摊 DEA 纳什讨价还价模型, *系统工程理论与实践*, 34(3), 756-768, 2014.
- [17] 刘业政, 叶俊杰, 蒋玮, 徐俊芬, 基于图上随机游动的 Web 页面布局可用性评价, *管理科学学报*, 16(9), 75-81, 2013.
- [18] 孙春华, 刘业政, 基于产品特征词关系识别的评论倾向性合成方法, *情报学报*, 32(8), 844-852, 2013.
- [19] 丁正平, 刘业政, 存在搭便车时双渠道的定价均衡及协调策略, *系统工程学报*, 28(3):



370-376, 2013

- [20] 陈思凤, **刘业政**, 混合状态变量和过程变量的风险评估模型, 控制与决策, 28(2), 309-312, 2013.
- [21] 张启平, **刘业政**, 刁翠霞, 面向社会经济发展水平的钢铁生产效率 DEA 动态评价, 系统工程理论与实践, 32(11), 2577-2584, 2012.
- [22] 陈思凤, **刘业政**, 姜元春, 于彪, 考虑决策风险偏好的自适应支持向量机模型, 系统仿真学报, 24(6), 1200-1206, 2012.
- [23] 刁翠霞, 陈思凤, **刘业政**, 基于 SVM 求解不均衡数据集分类的主观权重约束方法, 管理工程学报, 26(3), 146-150, 2012.
- [24] 王艳荣, **刘业政**, 农业产业集聚对产业增长贡献率的测度与分析[J], 中国农业科学, 45(15), 3197-3202, 2012.
- [25] 王艳荣, **刘业政**, 农业产业集聚形成机制的结构验证[J], 中国农村经济, 10, 77-85, 2012.
- [26] 孙春华, **刘业政**, 彭学仕, 一种含强度的基准词选择和词汇倾向性判别方法, 情报学报, 30(12), 1261-1267, 2011.
- [27] 何建民, 常传武, **刘业政**, 客户网上参与产品开发的“动机—行为”模型研究, 中国管理科学, 19(5), 173-181, 2011.
- [28] 林文龙, **刘业政**, 朱庆生, 奚冬芹, 基于混合隐 Markov 链浏览模型的 WEB 用户聚类与个性化推荐, 情报学报, 28(04), 557-564, 2009.
- [29] 凌海峰, **刘业政**, 杨善林, 基于蚁群算法与 K-means 算法相结合的 Web 用户聚类, 情报学报, 28(01), 105-108, 2009.
- [30] **刘业政**, 姜元春, 证据信度的效用分析, 系统工程理论与实践, 28(3), 103-110, 2008.
- [31] **刘业政**, 姜元春, 基于模糊距离和神经网络的自适应群决策方法, 系统工程学报, 23(1), 8-15, 2008.
- [32] **刘业政**, 林文龙, 可变多阶 Markov 链模型及在 WWW 个性化推荐中的应用, 情报学报, 27(6), 819-824, 2008.
- [33] 姜元春, **刘业政**, 林文龙, 刘晓, 基于粗糙集与证据理论的决策规则合成方法, 系统仿真学报, 20(04), 951-955, 2008.
- [34] **刘业政**, 徐德鹏, 姜元春, 多属性群决策中权重自适应调整的方法, 系统工程与电子技术, 01, 45-48, 2007.
- [35] **刘业政**, 林文龙, 焦宁, 姜元春, WEB 站点结构优化仿真, 系统仿真学报, 19(20), 4685-4688, 2007.
- [36] 杨善林, **刘业政**, 李亚飞, 基于 Rough Sets 理论的证据获取与合成方法[J], 管理科学学报, 8(5), 69-75, 2006.
- [37] 马溪骏, 凌海峰, **刘业政**, 等, 基于蚁群算法的群体用户兴趣导航路径发现[J], 中国管理科学, 14(3), 56-59, 2006.



- [38] **刘业政**, 张启平, 姜元春, 决策单元交叉效率的自适应群评价方法, 中国管理科学. (已录用)
- [39] **刘业政**, 杜亚楠, 姜元春, 杜非, 基于热度曲线分类建模的微博热门话题预测, 模式识别与人工智能. (已录用)
- [40] 何军, **刘业政**, 基于社交关系和影响力的在线社交网络用户兴趣偏好获取方法研究, 情报学报. (已录用)

(2) 会议论文

- [1] 王锦坤, 姜元春, **刘业政**, 面向实体超市的商品推荐系统架构, 信息系统协会中国分会第五届学术年会(CNAIS 2013), 2013/10/17-2013/10/19, 431-434, 贵阳, 2013
- [2] 陆安, **刘业政**, 基于连续影响函数的群体观点演化模型与仿真, 第十四届计算机模拟与信息技术学术(大会唯一一等奖论文), 2013.7, 昆明
- [3] **Yezheng Liu**, Zhengping Ding, Revenue Sharing Contract in Dual Channel Supply Chain in Case of Free Riding, 4th International Conference on Intelligent Decision Technologies (IDT), 2012/5/22-2012/5/25, pp 459-469, Gifu, 2012
- [4] 姬建睿, 姜元春, **刘业政**, 中文评论中产品特征与情感识别研究, 信息系统协会中国分会第四届学术年会(CNAIS 2011), 2011/12/2-2011/12/4, 上海, 2011
- [5] Wenlong Lin, **Yezheng Liu**, A Novel Website Structure Optimization Model for More Effective Web Navigation, First International Workshop on Knowledge Discovery and Data Mining, Melbourne, Australia 01, 36-41, 2008.
- [6] **Yezheng Liu**, Yuanchun Jiang, Xiao Liu, Shanlin Yang, Acquisition and Combination of Evidence Based on Rough Set Theory, International Conference on Industrial Engineering and Systems Management, Beijing, China, May 30-June 2, 2007.
- [7] Haifeng Ling, **Yezheng Liu**, Shanlin Yang, An Ant Colony Model for Dynamic Mining of Users Interest Navigation Patterns, 2007 IEEE International Conference on Control and Automation, 281-283, Guangzhou China, May 30-June 1, 2007
- [8] Yuan-Chun Jiang, **Ye-Zheng Liu**, Xiao Liu et al, Constructing Associative Classifier Using Rough Sets and Evidence Theory, Lecture Notes in Artificial Intelligence, 2007.5:263-271, 2007 Joint Rough Set Symposium MaRS Center, Toronto, Canada, May 14-16, 2007
- [9] **Yezheng Liu**, Yuanchun Jiang, Wenlong Lin, A Rough Set and Evidence Theory Based Method for Fraud Detection, IEEE Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA), Dalian, China, 6, 1538-1542, 2006.
- [10] Yuanchun Jiang, **Yezheng Liu**, An Attribute Reduction Method Based on Ant Colony Optimization, IEEE Proceedings of the 6th World Congress on Intelligent Control and Automation (WCICA), Dalian, China, 6, 3542-3546, 2006.
- [11] **Yezheng Liu**, Haifeng Ling, Shanlin Yang, An artificial ant colony methodology for users



navigation patterns mining, Proceedings of the 11th World Congress of International Fuzzy Systems Association, Beijing, China, 524-527, 2005.

- [12] Haifeng Ling, **Yezheng Liu**, Shanlin Yang, An Ant Colony Approach for Discovery of Users Preferred Navigation Paths, Proceedings of the 8th Pacific Asia Conference on Information Systems, Shanghai, China, 1654-1659, 2004.
- [13] Jian Hu, **Yezheng Liu**, Research on Business Intelligence in the Customer-Centric e-Business Environment, Proceedings of the 3rd international conference on e-commerce engineering, Zhe-Jiang University, International Academic Publishers, World Publishing Corporation, 484-488, 2003.

(3) 专著

- [1] **刘业政**, 姜元春, 张结魁, 网络消费者行为: 理论方法及应用, 科学出版社, 30 万字, 2011.

(4) 奖励

- [2] 杨善林, 周永务, 胡小建, 刘心报, 梁昌勇, **刘业政**, 朱卫东, 任明仑, 马溪骏, 面向企业的智能决策理论与方法, 国家教育部, 自然科学奖, 一等奖, 2008
- [3] 杨善林, 凌骥生, 褚伟, 任明仑, 左春荣, 梁昌勇, 左闯, **刘业政**, 蒋翠清, 李兴国, 金属液综合性能在线智能检测系统, 安徽省人民政府, 省科技进步奖, 一等奖, 2005
- [4] 杨善林, 孙自敏, 余本功, 徐东, **刘业政**, 任明仑, 马溪骏, 医院信息系统, 安徽省人民政府, 省科技进步奖, 二等奖, 2003
- [5] 高峰, 李晓钢, **刘业政**, 卜建民, 赵卫兵, 王立强, 安徽省建设信息系统(AHCISV1.0), 安徽省人民政府, 省科技进步奖, 三等奖, 2000
- [6] 杨善林, **刘业政**, 梁昌勇, 李兴国, 何建民, 任明仑, 倪志伟, 以创新人才培养为目标的教学团队建设, 国家教育部, 教学成果奖, 二等奖, 2014
- [7] 杨善林, **刘业政**, 梁昌勇, 李兴国, 赵惠芳, 任明仑, 何建民, 倪志伟, 朱卫东, 马溪骏, 创新型管理人才培养模式研究与实践, 国家教育部, 教学成果奖, 二等奖, 2009
- [8] 杨善林, **刘业政**, 马溪骏, 李兴国, 梁昌勇, 信息管理类专业教学研究与实践, 国家教育部, 教学成果奖, 二等奖, 2005
- [9] **刘业政**, 何建民, 任明仑, 梁昌勇, 倪志伟, 电子商务新专业建设研究与实践, 安徽省教育厅, 教学成果奖, 二等奖, 2010

(5) 专利

- [1] **刘业政**, 姜元春, 王锦坤等, 一种基于推荐概率融合的混合推荐方法, 申请号 201310637512.4.



2、杨善林，合肥工业大学，管理学院，硕士，教授，博士生导师，中国工程院院士。负责总体规划与研究应用方案设计。

2.1 个人简介

杨善林，中国工程院院士。主要从事智能决策理论与技术和信息管理理论与技术等方面的研究，并将研究成果运用于企业管理和社会管理中，在管理理论与管理实践的结合上取得了创新性成果。主持完成国家自然科学基金项目、国防基础研究项目、科技部支撑项目、国家“863”计划项目、国家发展和改革委员会项目、教育部博士点基金项目等纵向课题 20 多项和企业委托开发项目 20 余项；先后获国家科技进步二等奖 2 项，省部级科学技术一等奖 6 项、二等奖 3 项，获国家发明专利 3 项，软件著作权 6 项，撰写出版学术著作 5 部；近五年来，发表被 SCI 收录的学术论文 47 篇，被 EI 收录的学术论文 98 篇。现任“智能决策与信息系统”国家地方联合工程研究中心主任和“过程优化与智能决策”教育部重点实验室主任。主要学术兼职有：国务院学位委员会管理科学与工程学科评议组成员，国际信息系统学会中国分会副理事长，安徽省行为科学学会理事长，《预测》杂志主编。

2.2 受教育经历（从大学本科开始，按时间倒排序）

1988/09—1989/12，德国德累斯顿工业大学 CIMS 中心，访问学者
1986/07—1987/10，澳大利亚墨尔本大学，合作研究
1982/09—1985/01，合肥工业大学，计算机应用系，硕士
1978/09—1982/08，合肥工业大学，计算机应用系，学士

2.3 研究工作经历（按时间倒排序）

（1）工作经历

2013/11—至今，合肥工业大学管理学院，教授（中国工程院院士）
2006/07—2013/10，合肥工业大学管理学院，教授
2002/01—2006/06，合肥工业大学管理学院，副校长/教授
1996/07—2001/12，合肥工业大学管理学院，院长、副校长/教授
1994/01—1996/06，合肥工业大学管理学院，院长/副教授、教授
1985/02—1993/12，合肥工业大学计算机与信息学院，助教、讲师、副教授

（2）主持和参加的研究课题

- [1] 云计算环境下的多源信息服务系统研究，国家自然科学基金重点项目，2012.1-2016.12.
- [2] 云计算环境下的智能决策方法研究，国家自然科学基金项目，2011.1-2013.12.
- [3] 群决策理论与方法研究，国家自然科学基金重点项目，2007.1-2010.12.
- [4] 可信软件测试和可信性评估研究，国家自然科学基金重大研究计划项目，2008.1-2010.12.



- [5] “十二五”重大产业基地发展规划思路研究, 国家发改委委托项目, 2009.7-2010.7.
- [6] 江北产业集中区产业发展规划, 江北产业集中区国际招标项目, 2010.3-2010.11.
- [7] 面向供应链的制造工程协同优化与集成制造执行系统, 企业委托项目, 2005.1-2008.12.
- [8] 面向复杂决策任务基于 agent 组织的分布式智能决策支持系统研究, 国家自然科学基金项目, 2005.1-2007.12.
- [9] 集成定性推理理论的智能决策支持系统研究, 国家自然科学基金项目, 2002.1-2004.12.
- [10] 基于知识发现的综合决策支持系统研究, 国家自然科学基金项目, 2000.1-2002.12.

2.4 近五年发表的与本项目相关的主要论著

(1) 期刊论文

- [1] Fu Chao, Huhns Michael, **Yang Shanlin**, A consensus framework for multiple attribute group decision analysis in an evidential reasoning context, Information Fusion, 17(SI), 22-35, 2014.
- [2] Fu Chao, **Yang Shanlin**, Conjunctive combination of belief functions from dependent sources using positive and negative weight functions, Expert Systems With Applications, 41(4, 2), 1964-1972, 2014.
- [3] Pei Jun, Liu Xinbao, Panos M. Pardalos, Fan Wenjuan, **Yang Shanlin**, Wang Ling, Application of an effective modified gravitational search algorithm for the coordinated scheduling problem in a two-stage supply chain, International Journal of Advanced Manufacturing Technology, 70, (1-4), 335-348, 2014.
- [4] Hou Liqiang, **Yang Shanlin**, Volatility forecast based on intelligent EGARCH error correction model, Tehnicki Vjesnik-Technical Gazette, 20(6), 961-968, 2013.
- [5] Xu Dayu, **Yang Shanlin**, Liu Renping, A mixture of HMM, GA, and Elman network for load prediction in cloud-oriented data centers, Journal of Zhejiang University-Science C-Computers & Electronics, 14(11), 845-858, 2013.
- [6] Yu Shuhao, **Yang Shanlin**, Su Shoubao, Self-Adaptive Step Firefly Algorithm, Journal Of Applied Mathematics, 2013, 2013.
- [7] Zhou Kai-le, **Yang Shanlin**, Shen Chao, A review of electric load classification in smart grid environment, Renewable & Sustainable Energy Reviews, 24, 103-110, 2013.
- [8] Chen Yuwang, Yang JianBo, Xu DongLing, **Yang Shanlin**, On the inference and approximation properties of belief rule based systems, Information Sciences, 234, 121-135, 2013.
- [9] Zhang Chengtang, **Yang Shanlin**, Decision Analysis of Advertising and Price for Bilateral Competing Supply Chain, Mathematical Problems in Engineering, 2013.
- [10] He Yaoyao, **Yang Shanlin**, Xu Qifa, Short-term cascaded hydroelectric system scheduling based on chaotic particle swarm optimization using improved logistic map, Communications



- in Nonlinear Science and Numerical Simulation, 18(7),1746-1756, 2013.
- [11] **Yang Shanlin**, Bai Yu, Wang Sufeng, Feng Nanping, Evaluating the transformation of China's industrial development mode during 2000-2009, Renewable & Sustainable Energy Reviews, 20, 585-594, 2013.
- [12] Chen Zhiqiang, **Yang Shanlin**, Wang Xiaojia, PLS-SVR Optimized by PSO Algorithm for Electricity Consumption Forecasting, Applied Mathematics & Information Sciences, 7(1), 331-338, 2013.
- [13] Cheng Bayi, Wang Qi, **Yang Shanlin**, Hu Xiaoxuan, An improved ant colony optimization for scheduling identical parallel batching machines with arbitrary job sizes, Applied Soft Computing, 13(2), 765-772, 2013.
- [14] Chu Chengbin, Chu Feng, Zhong Jinhong, **Yang Shanlin**, A polynomial algorithm for a lot-sizing problem with backlogging, outsourcing and limited inventory, Computers & Industrial Engineering, 64(1), 200-210, 2013.
- [15] 李凯, 罗庆, 杨善林. 释放时间具有凸减函数约束的单机调度问题[J]. 系统工程理论与实践, 2013, 33(6): 1516-1522.
- [16] 张成堂, 杨善林. 双渠道回收下闭环供应链的定价与协调策略[J]. 计算机集成制造系统, 2013, 7: 028.
- [17] Cheng Bayi, **Yang Shanlin**, Ma Ying, Minimising makespan for two batch-processing machines with non-identical job sizes in job shop, International Journal of Systems Science, 43(12), 2185-2192, 2012.
- [18] Shi M-j, **Yang Shanlin**, Zhu Shixin, Good p-Ary Quasic-Cyclic codes from cyclic codes over $F_p + NU F_p$, Journal of Systems Science & Complexity, 25(2), 375-384, 2012.
- [19] Wang Xiaojia, Chen Zhiqiang, **Yang Shanlin**, Forecasting Modeling and Simulation Analysis of A Power System in China, Based on a Class of Semi-Parametric Regression Approach, South African Journal of Industrial Engineering, 23(3), 154-168, 2012.
- [20] Ding Shuai, Ma Xijun, **Yang Shanlin**. A software trustworthiness evaluation model using objective weight based evidential reasoning approach, Knowledge and Information System, 33(1),171-189, 2012.
- [21] Yang Jianbo, Xu Dongling, **Yang Shanlin**. Integrated efficiency and trade-off analyses using a DEA-oriented interactive minimax reference point approach, Computers & Operations Research, 39(5), 1062-1073, 2012.
- [22] Fu Chao, **Yang Shanlin**, The conjunctive combination of interval-valued belief structures from dependent sources, International Journal of Approximate Reasoning, 53(5), 69-785, 2012.
- [23] Fu Chao, **Yang Shanlin**, An evidential reasoning based consensus model for multiple



- attribute group decision analysis problems with interval-valued group consensus requirements, *European Journal of Operational Research*, 223(1), 167-176, 2012.
- [24] Fu Chao, **Yang Shanlin**, The combination of dependence-based interval-valued evidential reasoning approach with balanced scorecard for performance assessment, *Expert Systems With Applications*, 39(3), 3717-3730, 2012.
- [25] Cheng Bayi, **Yang Shanlin**, Hu Xiaoxuan, et al, Minimizing makespan and total completion time for parallel batch processing machines with non-identical job sizes, *Applied Mathematical Modelling*, 36(7), 3161-3167, 2012.
- [26] Fu Chao, **Yang Shanlin**, Group consensus based on evidential reasoning approach using interval-valued belief structures, *Knowledge-Based Systems*, 35, 201-210, 2012.
- [27] Luo He, **Yang Shanlin**, Hu Xiao-jian, et al, Agent oriented intelligent fault diagnosis system using evidence theory, *Expert Systems With Applications*, 39(3), 2524-2531, 2012.
- [28] Ding Shuai, **Yang Shanlin**, Fu Chao, A novel evidential reasoning based method for software trustworthiness evaluation under the uncertain and unreliable environment, *Expert Systems With Applications*, 39(3), 2700-2709, 2012.
- [29] Liu Xinbao, Pei Jun, Fan Wenjuan, **Yang Shanlin**. Optimization for batch scheduling problem based on mould capability, *International Journal of Advancements in Computing Technology*, 4(17): p 211-219. 2012.
- [30] 杨善林,黄志斌,任雪萍,工程管理中的辩证思维,中国工程科学,02,14-24, 2012.
- [31] 杨善林,罗贺,丁帅,基于云计算的多源信息服务系统研究综述,管理科学学报,05,83-96, 2012.
- [32] 马华伟,靳鹏,杨善林,时变车辆路径问题的启发式算法,系统工程学报,02,256-262, 2012.
- [33] 石莉,杨善林,马英,杨颖,一种新的组合权重集结方法及合理性评价研究,系统工程学报,04,481-491,2012.
- [34] 马英,杨善林,汤大为,带机器准备时间的同类机调度问题的启发式算法,系统工程理论与实践,09,2022-2030, 2012.
- [35] 徐达宇,杨善林,罗贺.云计算环境下多源信息资源管理方法,计算机集成制造系统,09,2028-2039,2012.
- [36] 杨善林,钟金宏,复杂产品开发工程管理的动态决策理论与方法,中国工程科学,12,25-40, 2012.
- [37] 胡笑旋,杨善林,华娟,基于图模型的反事实推理在决策后评估中的应用,中国管理科学,06,160-166,2012
- [38] Cheng F, **Yang Shanlin**, Akella R, et al, a Meta-Modelling Service Paradigm for Cloud Computing and ITs Implementation, *South African Journal of Industrial Engineering*, 22(2), 151-160, 2011.



- [39] Li Kai, **Yang Shanlin**, Ren Minglun, Single-machine scheduling problem with resource dependent release dates to minimise total resource-consumption, *International Journal of Systems Science*, 42(10), 1811-1820, 2011.
- [40] Wang Gang, Ma Jian, **Yang Shanlin**, Igf-Bagging: Information Gain Based Feature Selection for Bagging, *International Journal of Innovative Computing Info*, 7(11), 6247-6259, 2011.
- [41] Cheng F, **Yang Shanlin**, AKELLA R, et al, An Integrated Approach for Selection of Service Vendors in Service Supply Chain, *Journal of Industrial and Management Optimization*, 7(4), 907-925, 2011.
- [42] Li Kai, Shi Ye, **Yang Shanlin**, et al, Parallel machine scheduling problem to minimize the makespan with resource dependent processing times, *Applied Soft Computing*, 11(8), 5551-5557, 2011.
- [43] **Yang Shanlin**, Ma Ying, Xu Dongling, et al, Minimizing total completion time on a single machine with a flexible maintenance activity, *Computers & Operations Research*, 38(4), 755-770, 2011.
- [44] Li Kai, **Yang Shanlin**, Heuristic algorithms for scheduling on uniform parallel machines with heads and tails, *Journal of Systems Engineering and Electronics*, 22(3), 462-467, 2011.
- [45] Fu Chao, **Yang Shanlin**, Analyzing the applicability of Dempster's rule to the combination of interval-valued belief structures, *Expert Systems With Applications*, 38(4), 4291-4301, 2011.
- [46] Fu Chao, **Yang Shanlin**, An attribute weight based feedback model for multiple attributive group decision analysis problems with group consensus requirements in evidential reasoning context[J], *European Journal of Operational Research*, 212(1), 179-189, 2011.
- [47] Li Kai, **Yang Shanlin**, Ma Hua-wei, A simulated annealing approach to minimize the maximum lateness on uniform parallel machines, *Mathematical and Computer Modelling*, 53(5-6), 854-860, 2011.
- [48] 李聪,梁昌勇,**杨善林**,电子商务协同过滤稀疏性研究:一个分类视角,管理工程学报,01,94-101, 2011.
- [49] 柳春锋,李凯,**杨善林**,基于改进DP 算法的具有优先序的变速机最小化成本调度,系统工程理论与实践,02,276-282, 2011.
- [50] 丁帅,鲁付俊,**杨善林**,夏承遗,一种需求驱动的软件可信性评估及演化模型,计算机研究与发展,04,647-655, 2011.
- [51] 胡笑旋,**杨善林**,张强,支持群决策过程建模的层次影响图模型,系统工程理论与实践,07,1312-1319, 2011.
- [52] **杨善林**,程飞,杨昌辉,服务供应链的信息共享机制及绩效研究,中国工程科



- 学,08,80-86,2011.
- [53] 杨善林,马英,鲁付俊,带不可用时间段的单机调度问题的启发式算法,系统工程学报,04,500-506,2011.
- [54] 杨善林,刘心报,刘林,程浩,敏捷型生产过程优化和制造执行系统,系统工程理论与实践,S1,181-186,2011.
- [55] 杨颖,杨善林,马英,胡小建,改进的基于多依赖性的R&D 项目组合选择模型,系统工程学报,06,834-841,2011.
- [56] Liu Xinbao, Pei Feng, Yang Jianbo, **Yang Shanlin** et al, An MAGDM approach combining numerical values with uncertain linguistic information and its application in evaluation of r&d projects, International Journal of Computational Intelligence, 3(5), 575-589, 2010.
- [57] Wang Xiaojia, **Yang Shanlin**, Ding Jing, et al, Dynamic GM(1,1) model based on cubic spline for electricity consumption prediction in smart grid, China Communications, 7(4), 83-88, 2010.
- [58] Fu Chao, **Yang Shanlin**, The group consensus based evidential reasoning approach for multiple attributive group decision analysis, European Journal of Operational Research, 206(3), 601-608, 2010.
- [59] Jiang Yuanchun, Liu Yezheng, Liu Xiao, **Yang Shanlin**, et al. Integrating classification capability and reliability in associative classification: A β -stronger model, Expert Systems With Applications, 37(5), 3953-3961, 2010.
- [60] Gu Dongxiao, Liang Changyong, Li Xiongguo, **Yang ShanLin**, Intelligent Technique for Knowledge Reuse of Dental Medical Records Based on Case-Based Reasoning, Journal of Medical Systems, 34(2), 213-222, 2010.
- [61] 马英,左春荣,杨善林,带不可用时间段和恶化加工时间的单机调度,系统工程学报,03,371-378,2010.
- [62] 孙超平,杨善林,付超,基于证据推理的平衡计分卡及其在绩效评估中的应用,中国管理科学,03,75-82,2010.
- [63] 何建民,杨善林,刘赞,两阶段的制造商和代理商广告投资模型研究,系统工程学报,04,473-478,2010.
- [64] 杨善林,马华伟,顾铁军,时变条件下带时间窗车辆调度问题的模拟退火算法,运筹学学报,03,83-90,2010.
- [65] 杨善林,罗贺,胡小建,基于焦元相似度的证据理论合成规则,模式识别与人工智能,22(2),169-175, 2009.
- [66] 杨善林,丁帅,褚伟,一种基于效用和证据理论的可信软件评估方法,计算机研究与发展,48(7), 1152-1159, 2009.
- [67] 杨善林,丁帅,付超,考虑信息源相关的软件可信性评估模型,中国管理科学,



17(6),163-169,2009.

- [68] 杨善林,朱克毓,付超,卢广彦,基于元胞自动机的群决策从众行为仿真,系统工程理论与实践,29(09), 115-124, 2009.
- [69] 马英,杨善林,储诚斌,带不可用时间段的部分可续型单机最大完工时间调度,系统工程理论与实践, 29(4), 128-134, 2009.
- [70] Li Ying, Zheng Jiaqiang, **Yang Shanlin**, Multi-agent-based fuzzy scheduling for shop floor, International Journal of Advanced Manufacturing TE, 49(5-8), 689-695, 2009.
- [71] Li Kai, **Yang Shanlin**, Non-identical parallel-machine scheduling research with minimizing total weighted completion times: Models, relaxations and algorithms, Applied Mathematical Modelling, 33(4), 2145-2158, 2009.
- [72] **Yang Shanlin**, Fu Chao, Constructing confidence belief functions from one expert, Expert Systems With Applications, 36(4), 8537-8548, 2009.
- [73] Wu Jian, Sun Boliang, Liang Changyong, **Yang Shanlin**, et al. A linear programming model for determining ordered weighted averaging operator weights with maximal Yager's entropy, Computers & Industrial Engineering, 57(3), 742-747, 2009.
- [74] 陆青,梁昌勇,杨善林,面向多模态函数优化的自适应小生境遗传算法,模式识别与人工智能,(1), 91-100, 2009.
- [75] 付超,杨善林,罗贺,异源证据间的一致度分析,系统工程理论与实践,第29 卷,第5期, 166-174 页, 2009.
- [76] 李聪,梁昌勇,杨善林. 基于粗糙集的不完备信息系统空值估算方法. 计算机集成制造系统, 2009, 03: 604-608+617.

(2) 会议论文

- [1] Chen Zhiqiang, **Yang Shanlin**, Hou Liqiang, Based IGARCH error correction of the PLS-SVR short-term load forecasting, Proceedings of the 2012 International Conference on Electrical and Electronics Engineering,2012, p 199-207, 2014.
- [2] Wang Haijiang, **Yang Shanlin**, Mid-term load forecasting based on modified grey model, Proceedings of the 2013 International Conference on Mechatronics and Automatic Control Systems , 2013,p 169-175, 2014.
- [3] Zhao Guoxiang, **Yang Shanlin**, IT service incident management model decision based on ELECTRE, Proceedings of 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013, p 514-517, 2013
- [4] Yang Ying, **Yang Shanlin**, Yang Jian, A BSC-ANP model for evaluation of strategic fit of new product development projects,2013 25th Chinese Control and Decision Conference, 2013, p 1643-1647, 2013.
- [5] Hu Xianghong, **Yang Shanlin**, Hou Weiguo, Xie Zhangyong, Quality and skilled personnel



construction situation and development direction in China, Proceedings of 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 2013, p 497-500, 2013.

- [6] Cheng Fei, **Yang Shanlin**, Ma Xijun , Equilibrium conditions in service supply chain, 2011 International Conference on Advanced in Control Engineering and Information Science, 2011, p 5100-5104, 2011.
- [7] Cheng Fei, **Yang Shanlin**, Ma Xijun, Equilibrium conditions in service supply chain, 2011 International Conference on Advanced in Control Engineering and Information Science, 2011, p 5100-5104, 2011.
- [8] Lu Zhiping, **Yang Shanlin** , Process of complex group decision-making and its structural model of interactions, 2010 International Conference on Computer Design and Applications, 2010, p V3336-V3340, 2010.

(3) 专著

- [1] **杨善林**, 复杂产品开发工程管理理论与方法, 科学出版社, 86 万字, 2012.
- [2] **杨善林**, 周永务, 李凯等, 制造工程管理中的优化理论与方法, 科学出版社, 84 万字, 2012.
- [3] **杨善林**, 胡小建, 复杂决策任务的建模与求解方法, 科学出版社, 31 万字, 2007.
- [4] **杨善林**, 智能决策方法与智能决策支持系统, 科学出版社, 42.5 万字, 2005.
- [5] **杨善林**, 倪志伟, 机器学习与智能决策支持系统, 科学出版社, 47 万字, 2004.

(4) 奖励

- [1] 尹同耀、**杨善林**、陆建辉、鲁付俊、高立新、梁昌勇、刘心报、余本功、沈浩杰、王磊, 轿车整车自主开发系统的关键技术研究及其工程应用, 国家科技部, 国家科技进步奖, 二等奖, 2008
- [2] **杨善林**、凌骥生、褚伟、马溪骏、周永务、左闯、任明仑、刘心报、余本功、潘若愚, 铸铁材质参数液态在线智能检测与质量控制系统, 国家科技部, 国家科技进步奖, 二等奖, 2006
- [3] **杨善林**、周永务、胡小建、刘心报、梁昌勇、刘业政、朱卫东、任明仑、马溪骏, 面向企业的智能决策理论与方法, 国家教育部, 教育部自然科学奖, 一等奖, 2008
- [4] **杨善林**、余本功、胡小建、鲁付俊、高立新、任明仑、马英、杨颖、刘刚、吴卓亚, 面向供应链的制造工程协同优化与集成制造执行系统, 安徽省科技厅, 安徽省科学技术奖(科技进步类), 一等奖, 2012
- [5] **杨善林**、刘心报、李小宇、刘林、安涛、许安平、程浩、汤金春、毛征、裴凤, 离散制造业生产优化与执行系统, 安徽省科技厅, 安徽省科学技术奖(科技进步类), 一等奖, 2008
- [6] 尹同耀、**杨善林**、陆建辉、鲁付俊、高立新、梁昌勇、刘心报、余本功、沈浩杰、金德



全、王磊、胡小建、杨颖、罗贺、程飞，轿车整车产品开发系统平台、流程优化与工程管理系统，中国机械工业联合会，中国机械工业科技进步奖，一等奖，2007

- [7] 杨善林、凌骥生、褚伟、任明仑、左春荣、梁昌勇、左闯、刘业政、蒋翠清、李兴国，金属液综合性能在线智能检测系统，安徽省科技厅，安徽省科学技术奖(科技进步类)，一等奖，2005
- [8] 杨善林、陈福集、李兴国、王志龙、郑能靖、左春荣、马溪骏、姚君遗、梁昌勇，安庆石化总厂计算机辅助管理与生产过程监测系统，安徽省科技厅，安徽省科技进步奖，一等奖，1996
- [9] 杨善林、周永务、刘心报、黄梯云、朱卫东、梁昌勇、任明仑、马溪骏，基于知识发现与定性推理的智能决策支持系统，安徽省科技厅，安徽省科学技术奖(自然科学类)，二等奖，2004
- [10] 杨善林、孙自敏、徐冬、刘业政、余本功、任明仑、马溪骏，医院信息系统，安徽省科技厅，安徽省科学技术奖(科技进步类)，二等奖，2003
- [11] 杨善林、洪斌、马溪骏、王孝武，大型冷冻站制冷控制仿真培训系统，安徽省科技厅，安徽省科技进步奖，二等奖，1996

(5) 专利

- [1] 杨善林，凌骥生，褚伟等，金属液综合性能在线检测智能检测系统，授权号 ZL2006 1 0002850.0.
- [2] 杨善林，凌骥生，褚伟等，金属液综合性能在线检测用传感器，授权号 ZL2006 1 0002851.5.
- [3] 杨善林，丁静，罗贺等，云计算环境下的分类规则挖掘方法，授权号 2012 1 0203816.5.
- [4] 杨善林，范雯娟，罗贺等，基于调度目标空间的云计算资源调度方法，申请号 201210194890.5.
- [5] 杨善林，王晓佳，杨昌辉等，一种基于 Markov 链的正交化插值的 GM(1,1)模型预测方法，申请号 201310025028.6.
- [6] 杨善林，王晓佳，杨昌辉等，一种基于组合插值的 GM(1, 1)模型预测方法，申请号 201310025197.X.
- [7] 杨善林，王晓佳，杨昌辉等，一种基于三次样条的 GM(1, 1)模型预测方法，申请号 201310025218.8.
- [8] 杨善林，罗贺，王洪波等，一种基于流形的 AHP 云决策方法，申请号 201210184939.9.
- [9] 杨善林、王国强、罗贺等，基于 VR-Forces 仿真平台的多无人机协同任务规划仿真系统，申请号 201310751055.1.
- [10] 徐达宇，杨善林，章政等，基于自适应弹性控制的云计算资源优化配置方法，申请号 201310095506.0.
- [11] 徐达宇，杨善林，章政等，一种新的公有云架构及其虚拟化资源自适应配置方法，申请



号 201310245199.X.

[12] 杨善林, 丁帅, 罗贺等, 一种基于多属性匹配的云服务推荐方法及装置, 申请号 201110377164.2.

[13] 杨善林, 胡小建, 任明仑, 一种基于数据挖掘的大型装备 MRO 知识获取方法, 申请号 201110335215.5.

3、倪志伟, 合肥工业大学, 管理学院, 博士, 教授, 博士生导师。负责“专题二: 面向商务管理的大数据分析与计算方法”研究。

3.1 个人简介

倪志伟教授是中国离散系统仿真协会专家委员会委员, 安徽省软件工程学会专家组副组长, 安徽省物流协会标准委员会专家组成员, 合肥工业大学国家教育部“过程优化和智能决策”重点实验室学术委员会成员, 国家教育部管理学科创新团队的主要成员。倪志伟教授的主要研究领域包括人工智能、决策支持理论和技术、机器学习和智能软件系统。曾主持或参加 20 多项国家及省部级课题, 2010 年主持完成了国家“863”项目“面向制造业售后服务的商务智能关键技术的研究”, 2011 年主持完成了国家自然科学基金项目“商务智能中的动态数据挖掘与分形技术研究”; 作为主要人员, 参加国家“863”计划的主题项目“集团企业云制造服务平台共性关键技术研究”。2009 年获得国家级教学成果二等奖 1 项, 2008 年获得安徽省教学成果特等奖 1 项, 获得省级优秀学术论文三等奖 2 项、二等奖 1 项, 省科技进步三等奖 1 项, 省科技进步二等奖 2 项。近年来, 公开发表与本项目相关学术论文 60 多篇, 其中 SCI/EI 检索论文近 30 篇。

3.2 受教育经历(从大学本科开始, 按时间倒排序)

1999/09—2002/06, 中国科学技术大学, 信息技术学院, 博士

1988/09—1991/06, 安徽大学, 计算机系, 硕士

1980/09—1984/06, 安徽大学, 计算机系, 学士

3.3 研究工作经历(按时间倒排序)

(1) 工作经历

2002/09—至今, 合肥工业大学, 管理学院, 教授/博导, 所长

2003/11—2004/03, 澳大利亚斯维本科技大学, 信息技术研究中心, 高级访问学者

2000/08—2002/09, 安徽大学, 电子学院, 副教授, 副院长

1995/05—2000/08, 安徽大学, 计算机系, 副教授, 系副主任

1992/07—1995/05, 安徽大学, 计算机系, 讲师

1984/04—1992/07, 安徽大学, 计算机系, 助教



(2) 主持和参加的研究课题

- [1] 云计算环境下基于分形数据挖掘技术的商务智能系统的研究，国家自然科学基金，2013.01 -2016.12，主持
- [2] 集团企业云制造服务平台共性关键技术研究，国家“863”计划主题项目，2011.01 -2013.03，子课题主持人
- [3] 商务智能中的动态数据挖掘与分形技术研究，国家自然科学基金，2009.01-2011.12，主持
- [4] 面向制造业售后服务的商务智能关键技术的研究，国家“863”计划项目，2007.10-2010.06，主持
- [5] 群决策理论与方法研究，国家自然科学基金重点项目，2007-2010，参加者
- [6] RFID 技术在汽车制造过程与质量跟踪系统中的应用，国家“863”计划项目，2007.10-2010.10，参加者
- [7] 面向隐性目标决策问题的智能决策方法与支持系统研究，国家自然科学基金项目，2008.01-2011.12，参加者
- [8] 安徽省交通管理信息系统，安徽省交通厅委托的横向项目，2012.01-2013.06，主持

3.4 近五年发表的与本项目相关的主要论著

(1) 期刊论文

- [1] Tang Liyang, **Ni Zhiwei**, Xiong Hui, et al. Locating targets through mention in Twitter[J], World Wide Web, 1-31, 2014.
- [2] Gu L, **Ni Zhiwei**, A Similarity Measures Algorithm for CBR Based on Matrix Iterative Learning[J], International Journal of Applied Mathematics and Statistics™, 43(13), 271-280, 2013.
- [3] **Ni Zhiwei**, Gao Yazhuo, Hu Tanglei, Statistical Modeling for OLAP Approximate Query, Journal of computational information systems, 8(9), 3605–3612, 2012.
- [4] Zhang Yiwen, **Ni Zhiwei**, Jin Ying, et al, Large-scale Partner Selection of Virtual Enterprise on the Cloud Manufacturing Service Platform[J], Journal of Information & Computational Science, 9(14), 3995-4006, 2012.
- [5] **Ni Zhiwei**, Wu Zhangjun, Support Vector Machine Based Knowledge Discovery in CBR System[J], International Review on Computers & Software, 7(1), 2012.
- [6] **Ni Zhiwei**, Guo Junfeng, Wang Li, Gao Yazhuo, An Efficient Method for Improving Query Efficiency in Data Warehouse, Journal of software, 6(5), 857-865, 2011.
- [7] Wu Zhangjun, Liu Xiao, **Ni Zhiwei**, Yuan Dong, Yang Yun, A Market-Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems, the Journal of Supercomputing, 63(1), 256-293, 2013.
- [8] Ni LiPing, **Ni ZhiWei**, Gao YaZhuo, Stock trend Prediction Based on Fractal Feature



- Selection and Support Vector Machine, Expert systems with applications, 38, 5569-5576, 2011.
- [9] Liu Xiao, **Ni Zhiwei**, et al, A Novel Statistical Time-Series Pattern based Interval Forecasting Strategy for Activity Durations in Workflow Systems, Journal of Software and System, 84, 354-376, 2011.
- [10] Liu Xiao, **Ni Zhiwei**, Chen J, et al. A probabilistic strategy for temporal constraint management in scientific workflow systems[J], Concurrency and Computation: Practice and Experience, 23(16), 1893-1919, 2011.
- [11] Liu Xiao, **Ni Zhiwei**, Wu Zhangjun, et al. A novel general framework for automatic and cost-effective handling of recoverable temporal violations in scientific workflow systems[J], Journal of Systems and Software, 84(3), 492-509, 2011.
- [12] **倪志伟**, 肖宏旺, 伍章俊等, 基于改进离散型萤火虫群优化算法和分形维数的属性选择方法[J], 模式识别与人工智能, 26(12), 1169-1178, 2013.
- [13] 伍章俊, 刘晓, **倪志伟**, 基于混沌时间序列的云 workflow 活动运行时间预测模型[J], 计算机集成制造系统, 19(8), 1920-1927, 2013.
- [14] **倪志伟**, 张敞, 高雅卓等, 基于变量选择的 OLAP 查询推荐仿真模型[J], 系统仿真学报, 25(011), 2534-2539, 2013.
- [15] 倪丽萍, **倪志伟**, 一种基于趋势分形维数的股指时间序列相似性分析方法, 系统工程理论与实践, 9, 1900-1907, 2012.
- [16] 叶红云, **倪志伟**, 倪丽萍, 一种检测兴趣漂移的图结构推荐系统[J], 小型微型计算机系统, 33(4), 700-706, 2012.
- [17] 吴昊, **倪志伟**, 王会颖, 基于 MapReduce 的蚁群算法[J], 计算机集成制造系统, 18(07), 1503-1509, 2012.
- [18] 高雅卓, **倪志伟**, 倪丽萍, 连续属性上的 OLAP 查询建模方法研究, 情报学报, 30(30), 372-379, 2011.
- [19] **倪志伟**, 公维峰, 周之强, 等. 数据流中随机型分形维数计算方法研究[J]. 计算机科学, 38(4), 209-212, 2011.
- [20] **倪志伟**, 吴昊等, 基于改进的经验模态分解的时间序列匹配算法, 系统仿真学报, 23(11), 2395-2399, 2011.
- [21] 唐李洋, **倪志伟**, 李应, 基于 Cassandra 的可扩展分布式反向索引的构建[J], 计算机科学, 38(6), 187-190, 2011.
- [22] 叶红云, **倪志伟**, 倪丽萍, 基于 ACV 约束的序列模式挖掘算法, 模式识别与人工智能, 23(6), 802-808, 2010.
- [23] **倪志伟**, 姜苗, 王超, 戴奇波, 在线挖掘数据流混合窗口中闭频繁项集, 系统仿真学报, 9, 19-23, 2010.



- [24] 刘慧婷, 倪志伟, 基于 EMD 与 K-means 算法的时间序列聚类, 模式识别与人工智能, 22(5), 803-808, 2009.
- [25] 倪丽萍, 倪志伟, 吴昊, 叶红云, 基于分形维数和蚁群算法的属性选择方法, 模式识别与人工智能, 22(2), 293-298, 2009.
- [26] 郭峻峰, 倪志伟, 高雅卓, 伍章俊, 一种提高数据仓库查询效率的有效方法, 计算机集成制造系统, 15(12), 2451-2457, 2009.
- [27] 吴姗, 倪志伟, 罗贺, 郑盈盈, 一种基于密度的无监督联系发现方法, 中国管理科学, 16, 29-32, 2008.

(2) 会议论文

- [1] Gao Yazhuo, Ni Zhiwei, Ni Liping, Compression of OLAP Cubes for Aggregate Query Based on Copula Method, Proceedings of 2010 International Conference on Business Intelligence and Financial Engineering, 67-71, 2010.
- [2] Wu Zhangjun, Ni Zhiwei, Gu Lichuan, Liu Xiao, A Revised Discrete Particle Swarm Optimization for Cloud Workflow Scheduling, 2010 International Conference on Computational Intelligence and Security (CIS2010), 184-188, 2010.
- [3] Liu Xiao, Ni Zhiwei, Wu Zhangjun, Yuan Dong, Chen Jinjun, Yang Yun, A Framework for Handling Fine-Grained Recoverable Temporal Violations in Scientific Workflows, 16th IEEE International Conference on Parallel and Distributed Systems (ICPADS2010), 7-11, 2010.

(3) 专著

- [1] 倪志伟, 倪丽萍, 刘慧婷, 贾瑞玉, 动态数据挖掘, 科学出版社, 32 万字, 2010.
- [2] 倪志伟, 李锋刚, 毛雪岷, 智能管理技术与方法, 科学出版社, 35 万字, 2007.

(4) 奖励

- [1] 倪志伟 (5), 基于 GIS 技术的安徽省重大农业气象灾害测评系统, 安徽省人民政府, 省科技进步奖, 二等奖, 2003
- [2] 倪志伟 (2), 安徽省公路建设工程动态管理系统, 安徽省人民政府, 省科技进步奖, 三等奖, 2009

(5) 专利

- [3] 倪志伟, 基于数据流的车辆故障诊断系统, 专利号: ZL 2010 0183798.X, 授权公告日: 2012 年 2 月 29 日

4、李军, 西南交通大学, 经济管理学院, 博士, 教授, 博士生导师。负责“专题一: 面向商务管理的大数据获取方法与可视化技术”研究。

4.1 个人简介



李军教授现任西南交通大学经济管理学院党委书记，研究领域包括物流与供应链管理，博弈理论及应用，优化与决策。先后主持国家自然科学基金 5 项、国家社会科学基金 1 项、博士点基金 1 项、教育部人文社科项目 1 项、欧盟项目等其他项目 10 余项。在国内外高水平期刊发表学术论文 100 余篇，出版专著四部，主、参编教材 3 部。入选教育部新世纪优秀人才，评为四川省学术技术带头人、四川省教学名师，获得成都市“十大杰出”青年称号。获得霍英东教育基金会青年教师奖（教学类）一等奖，詹天佑西南交通大学（管理奖），四川省科技进步二等奖 2 项，四川省哲学社会科学优秀成果三等奖 1 项，铁道部优秀教材一、二等奖各 1 项。

4.2 受教育经历（从大学本科开始，按时间倒排序）

2004/10—2005/10，香港中文大学，系统工程与工程管理系，博士后研究
1996/09—2000/04，西南交通大学，经济管理学院管理科学与工程系，博士
1992/05—1994/05，西南交通大学，经济管理学院管理科学与工程系，硕士
1985/09—1989/07，西南交通大学，工业管理工程专业，学士

4.3 研究工作经历（按时间倒排序）

（1）工作经历

2001/07— 至今， 西南交通大学，经济管理学院，教授，博导
2000/07— 2001/06，西南交通大学，经济管理学院，教授
1996/09— 2000/04，西南交通大学，经济管理学院，副教授
1994/06— 1996/05，西南交通大学，经济管理学院，讲师
1992/05— 1994/05，西南交通大学，经济管理学院，助教

（2）主持和参加的研究课题

- [1] 具有竞争行为的公路合作运输决策研究，国家自然科学基金，2013.01-2016.12，主持
- [2] 面向乘客出行行为的公交客流分配及应用研究，教育部人文社会科学研究规划基金项目，2012.01-2014.12，主持
- [3] 基于文字信息量与短期记忆的交通指路标志视认性研究，四川省哲学社会科学重点研究基地应用心理学研究项目，2014.01-2015.12，主持
- [4] 服务运作与项目风险管理决策研究，中央高校基本科研业务费专项资金，2010.01-2012.12，主持
- [5] 行为决策理论及其在管理中的应用研究，教育部创新团队，2009.01-2011.12，主研
- [6] 震后交通恢复与物资调运决策研究，国家自然科学基金应急项目，2008.08-2009.02
- [7] 基于 EPR 的循环经济决策问题研究，国家社会科学基金，2007.09—2009.08，主持
- [8] 新世纪优秀人才支持计划，教育部，2005.01-2007.12，主持
- [9] 具有时变特性的有害物品运输路径研究，国家自然科学基金，2005.01-2007.12，主持
- [10] 第三方存贮一路径问题研究，国家自然科学基金，2003.01-2005.12，主持



- [11] 多层结构的时间窗存贮一路径问题研究, 高等学校博士学科点专项科研基金资助课题, 2004.01-2006.12, 主持
- [12] 生产者责任延伸制度下企业逆向物流研究, 四川省哲学社会科学“十一五”规划重点项目, 2007.01-2008.12, 主持
- [13] International Logistics Operation Managemen, European Commission Asia-Link Programme, 2006.01-2008.12, 主持
- [14] 货运车辆调度优化理论及应用研究, 国家自然科学基金, 1998.01-2000.12, 主持
- [15] 不确定信息条件下动态车辆路径问题研究, 国家自然科学基金, 2001.01-2003.12, 主研
- [16] 城市货运汽车的科学调度, 国家自然科学基金, 1991.01-1992.12, 主研

4.4 近五年发表的与本项目相关的主要论著

(1) 期刊论文

- [1] **Jun Li**, Hairong Feng, Yinlian Zeng, Inventory games with permissible delay in payments, European Journal of Operational Research, 234(3), 694 -700, 2014.
- [2] **Jun Li**, Hairong Feng, Kun-Jen Chung, Using the Algebraic Approach to Determine the Replenishment Optimal Policy with Defective Products, backlog and delay of payments in the Supply Chain Management, Journal of Industrial and Management Optimization, 8(1), 263-269, 2012.
- [3] **李军**, 岳青, 冯海荣, 易腐品联合订购博弈分析, 西南交通大学学报, 47(6), 1049-1056, 2012.
- [4] **Jun Li**, Hairong Feng, Retailer's optimal replenishment policy under cash discount and trade credit linked to order quantity. International Journal of Information and Decision Sciences, 2(4), 375-400, 2010.
- [5] **李军**, 魏洁, 基于 EPR 制度的逆向物流研究与应用综述, 软科学, 24(4), 43-47, 2010.
- [6] **李军**, 蔡小强, 易腐性产品运输设施选择博弈, 管理科学学报, 12(1), 28-37, 2009.
- [7] **Jun Li**, Jiongwei Mao, An inventory model of perishable item with two types of retailers, Journal of the Chinese Institute of Industrial Engineers, 26(3), 176-183, 2009.
- [8] **Jun Li**, Hairong Feng, Minchao Wang, A replenishment policy with defective products, backlog and delay of payments, Journal of Industrial and Management Optimization, 5(4), 867-880, 2009.
- [9] 赵达, **李军**, 马丹祥, 李妍峰, 随机需求库存-路径问题最优策略及其算法, 管理科学学报, 17(5), 14-24, 2014.
- [10] 赵达, **李军**, 马丹祥, 李妍峰, 直接配送下随机需求库存-路径问题最优平稳策略及其算法, 中国管理科学, 22(6), 71-78, 2014.
- [11] 赵达, **李军**, 马丹祥, 李妍峰, 求解硬时间窗约束下随机需求库存-路径问题的优化算法,



- 运筹与管理, 23(1), 26-32+38, 2014.
- [12] 林雪梅, **李军**, 周杰, 医院 CT 室病人到达规律分析, 工业工程与管理, 19(1), 103-107, 2014.
- [13] 郭鹏, 萧柏春, **李军**, 收益管理系统多分布无约束估计方法仿真比较, 计算机仿真, 31(2), 264-270, 2014.
- [14] Hairong Feng, **Jun Li**, Retailer's optimal replenishment and payment policies in the EPQ model under cash discount and two-level trade credit policy, Applied Mathematical Modelling, 37, 3322-3339, 2013.
- [15] 曾银莲, **李军**, 基于合作博弈的排队系统合作费用分配, 系统工程, 31(12), 103-108, 2013.
- [16] 郭鹏, 萧柏春, **李军**, 基于多需求分布的收益管理多舱位无约束估计模型, 系统科学与数学, 33(10), 1164-1177, 2013.
- [17] 冯海荣, **李军**, 曾银莲, 延期支付下的易腐品联合采购费用分配, 系统工程理论与实践, 33(6), 1411-1423, 2013.
- [18] 李妍峰, 高自友, **李军**, 基于实时交通信息的城市动态网络车辆路径优化问题, 系统工程理论与实践, 33(7), 1813-1819, 2013.
- [19] 曾鸷, **李军**, 朱晖, 实时信息下的乘客路径选择行为研究, 计算机应用, 33(10), 2964-2968, 2013.
- [20] 曾鸷, **李军**, 朱晖, 换乘行为影响下的城市公交配流算法, 交通运输工程学报, 13(4), 70-78, 2013.
- [21] 曾鸷, **李军**, 朱晖, 考虑乘客感知的公交随机用户均衡配流研究, 计算机应用, 33(4), 1149-1152+1168, 2013.
- [22] 曾鸷, **李军**, 合作博弈视角下城市道路交通拥堵收费研究, 运筹与管理, 22(1), 9-14, 2013.
- [23] P Guo, B C Xiao, **Jun Li**, Unconstraining Estimation in Revenue Management System: Research Overview and Prospects, Advances in Operations Research, 1-23, 2012.
- [24] 李妍峰, **李军**, 高自友, 动态规划启发式算法求解时变车辆调度问题, 系统工程理论与实践, 32(8), 1712-1718, 2012.
- [25] 李妍峰, **李军**, 高自友, 基于大规模邻域搜索技术的智能优化方法求解时变网络车辆调度问题, 管理科学学报, 15(1), 22-32, 2012.
- [26] 冯海荣, **李军**, 曾银莲, 易腐品供应链横向企业联合采购与费用分配研究, 系统科学与数学, 31(11), 1454-1466, 2011.
- [27] 郭鹏, 萧柏春, **李军**, 收益管理中非正态无约束估计的 EM 算法研究, 数理统计与管理, 30(6), 1077-1088, 2011.
- [28] 李妍峰, **李军**, 高自友, 时变网络环境下旅行商问题研究, 系统工程学报, 25(5),



585-591, 2010.

- [29] 陈六新, **李军**, 价格和库存均影响需求、部分拖后供给的易变质品的生产库存, 数学的实践与认识, 39(8), 61-67, 2009.
- [30] Baichun Xiao, Wei Yang, **Jun Li**, Optimal Reserve Price for the Generalized Second-Price Auction in Sponsored Search Advertising, Journal of Electronic Commerce Research, 10(3), 114-129, 2009.
- [31] 李妍峰, **李军**, 赵达, 动态搜索算法求解时间依赖型旅行商问题研究, 控制与决策, 24(2), 274-278, 2009.
- [32] 谢天帅, 赵玉双, **李军**, 第三方物流服务商道德风险的防范, 系统管理学报, 18(2), 137-141, 2009.
- [33] 郭晓林, **李军**, 有害物品运输风险度量模型的两种改进分析, 软科学, 23(8), 41-44, 2009.
- [34] 陈六新, **李军**, 谢天帅, 价格和库存均影响需求的易腐品模型, 商业研究, 5, 39-43, 2009.
- [35] 李良, 郭强, **李军**, 震后紧急搜救资源配置, 系统工程, 27(8), 1-7, 2009.
- [36] 郭鹏, 萧柏春, **李军**, 收益管理中考虑顾客策略行为的多航班无约束估计方法, 系统工程理论与实践 (录用)
- [37] 郭鹏, 萧柏春, **李军**, 收益管理中非正态无约束估计的 PD 方法, 数理统计与管理 (录用)
- [38] 曾银莲, **李军**, 刘云霞, 基于最大一致集的合作运输联盟稳定性分析, 系统科学与数学 (录用)
- [39] 赵达, **李军**, 马丹祥, 李妍峰, 基于固定分区策略下随机需求库存-路径问题算法, 管理科学学报 (录用)

(2) 会议论文

- [1] Ying Zeng, **Jun Li**, Zhu Hui, Transit Transfer Optimization with Capacity Constraint Consideration, Proceedings of 2013 International Conference on Management Science & Engineering (20th), 2013.
- [2] Yanfeng Li, Ziyao Gao, **Jun Li**, Real-time Vehicle Routing and Scheduling in Dynamic Urban Traffic Network, Proceedings of the 8th International Conference on Service Systems and Service Management, 1257-1262, 2011.
- [3] Yinlian Zeng, **Jun Li**, Xiaoqiang Cai, Economic Lot-sizing Games with Perishable Inventory, Proceedings of the 8th International Conference on Service Systems and Service Management, 1252-1256, 2011.
- [4] Hairong Feng, **Jun Li**, Ying Zeng, Retailer's optimal replenishment policy with defective products under cash discount and trade credit, The International Conference E-Product, E-Service and E-Entertainment, 4-6, 2010.
- [5] Ying Wang, **Jun Li**, Comparative Research on the Different Bargaining Power of



Competitive Manufacturers in Supply Chain, ICLEM 2010, 2010.

- [6] Mingchao Wang, **Jun Li**, The revenue sharing contract with two substitutable products: coordination and limitations, The 6th International Symposium on Management Engineering, 5-7, 2009.

(3) 著作

- [1] **李军**, 郭晓林, 魏航, 李继兵, 有害物品运输问题研究, 科学出版社, 27 万字, 2010.
- [2] **李军**, 谢天帅, 第三方物流服务定价方法, 科学出版社, 19.6 万字, 2010.
- [3] 赵昌文主编, **李军**, 李成金, 唐春勇, 李良, 郭强, 应急管理与灾后重建: 5.12 汶川特大地震若干问题研究: 《震后交通恢复与物资调运决策研究》第 4 章, 科学出版社, 67 万字, 2011.
- [4] **Jun Li**, Naiyi Ye, Peng Guo, Dongfang Electric Corporation Case Study: Virtual Enterprise in Contract and Services Project, International operations management, lessons in Global Business, Edited by Alberto F. De Toni, R.D. Franco, **J. Li**, Y. Li, G. Nassimbeni, M. Sartor, X. Zhao, X. Xu, Published by Gower Publishing Limited, 2011

(4) 奖励

- [1] **李军**, “易变质品与缺陷产品的合作运输与库存控制问题研究系列论文”, 四川省第十四次哲学社会科学优秀成果三等奖, 2010
- [2] **李军**, “不确定情景下路径选择理论与应用研究”, 四川省科技进步二等奖, 2008
- [3] **李军**, “不确定情景下路径选择理论与应用研究”, 成都市科技进步三等奖, 2008
- [4] **李军**, “城市货运卡车集中优化调度研究”, 四川省科技部, 四川省科技进步二等奖, 1995

5、何建民, 合肥工业大学, 管理学院, 博士, 教授, 硕士生导师。负责“专题三: 基于网络大数据的商业情报分析方法”研究。

5.1 个人简介

何建民教授现任合肥工业大学管理学院电子商务系主任, 研究领域和方向包括电子商务、商务智能系统、企业管理及其信息化。1984 年安徽大学计算机软件专业本科毕业, 此后一直在企事业单位和政府部门从事计算机软件系统的研发与技术管理工作, 先后担任过部门经理、副总经理、总工程师等职务。1999 年开始在合肥工业大学管理学院从事教学和科研工作, 同时考取在职研究生, 并获得管理学博士学位。近年来, 先后主持并完成了国家、省(部)、市级纵向和横向科研项目 10 余项, 并获得国家教学成果二等奖 2 项、省级教学成果特等奖 2 项、省社科成果二等奖 1 项、省科技成果三等奖 1 项; 在国外 SCI 检索期刊和国内著名核心期刊上发表学术论文 30 余篇。主编出版了高等教育国家级规划教材 3 部。



5.2 受教育经历（从大学本科开始，按时间倒排序）

2006/09—2010/06，合肥工业大学，管理学院，博士

1999/09—2002/06，合肥工业大学，管理学院，硕士

1980/09—1984/06，安徽大学，计算机系，学士

5.3 研究工作经历（按时间倒排序）

（1）工作经历

2010/12—至今，合肥工业大学管理学院，教授/硕导，系主任

2003/01—2010/11，合肥工业大学管理学院，副教授/硕导，系主任

1999/03—2002/12，合肥工业大学管理学院，高级工程师/硕导，系副主任

1996/03—1999/02，安徽省对外经济贸易委员会，高级工程师，总工程师

1984/07—1996/03，中国软件技术公司安徽分公司，工程师，副总经理

（2）主持和参加的研究课题

- [1] 面向互联网大数据的舆论领袖话语影响力评估和治理方法研究，教育部人文社科规划基金项目，2014.01-2016.12，主持
- [2] 网络社区舆情及舆论领袖识别方法研究，教育部人文社科规划基金项目，2011.01-2013.10，主持
- [3] 合芜蚌自主创新综合试验区创新平台建设及其运行机制研究，科技部软科学重大计划专项项目，2011.03-2012.03，主持
- [4] 云计算环境下基于分形数据挖掘技术的商务智能系统的研究，国家自然科学基金项目，2013.01-2016.12，主要参与者
- [5] 群决策理论与方法研究，国家自然科学基金重点项目，2007.01-2010.12，主要参与者
- [6] 面向制造业售后服务的商务智能关键技术研究，国家 863 重点项目，2007.01-2010.12，主要参与者
- [7] 安徽***公司市场营销与快餐直营连锁经营管理系统开发，企业委托横向课题，2012.01-2013.12，主持

5.4 近五年发表的与本项目相关的主要论著

（1）期刊论文

- [1] **Jianmin He**, Mengna Hu, Mingguang Shi, Yezheng Liu, Research on the Measure Method of Complaint Theme Influence on Online Social Network, Expert Systems with Applications, 41(13), 6039-6046, 2014.
- [2] **何建民**, 金加卫, 网络社区顾客抱怨影响力三维度量方法, 情报学报, 32(4), 421-427, 2013.
- [3] **何建民**, 贺涛, 基于 QFD 的网络社区话题影响力测评方法, 情报科学, 31(263), 96-99,



2013.

- [4] 何建民, 常传武, 刘业政, 客户网上参与产品开发的“动机-行为”模型研究, 中国管理科学, 19(5), 173-181, 2011.
- [5] 何建民, 吴锐侠, 杨善林, OSN下的客户领袖声音在线识别方法研究, 系统管理学报, 18(6), 692-699, 2009.
- [6] 何建民, 杨善林, JAC 聆听客户声音变革企业管理体系的实证研究, 汽车工程, 32(8), 737-742, 2010.
- [7] 何建民, 杨善林, 两阶段的制造商和代理商广告投资模型研究, 系统工程学报, 25(4), 473-478, 2010.
- [8] 江玮幡, 何建民, 基于粒子群优化的有向物流的车辆路径问题, 系统管理学报, 17(4), 475-479, 2008.
- [9] 何建民, 张义, 基于类熵距离测量的热点话题识别方法研究, 情报科学, 30(252), 1147-1150, 2012.
- [10] 何建民, 贺涛, 网络社区有影响力话题度量识别方法研究, 图书情报工作, 56(2), 195-200, 2012.
- [11] 何建民, 孙庆庆, 刘业政, 复合参数异步检测网络热点话题方法研究, 情报杂志, 30(10), 145-150, 2011.
- [12] 何建民, 刘明星, 面向特定主题的客户抱怨文本分类识别方法, 合肥工业大学学报(自然科学版), 33(12), 1884-1889, 2010.
- [13] 何建民, 徐艳玲, 网络环境下客户抱怨的贝叶斯分析及处理方法研究, 合肥工业大学学报(自然科学版), 33(6), 929-933, 2010.

(2) 会议论文

- [1] Jianmin He, Cheng Fu and Shanlin Yang, Research on the Design Method of Business-Orient Component-Based CRM System, Proceedings of the 7th World Congress on Intelligent Control and Automation, IEEE, 8298-8302, 2008.
- [2] 常传武, 何建民, 引导因特网客户参与产品开发的策略架构研究, 第八届全国电子商务教育与学术大会论文集, 合肥工业大学出版社, 112 – 117, 009.
- [3] 何建民, 陈志龙, 受在线产品评论影响的零售商与消费者主从微分博弈, 《第九届全国高校电子商务教育与学术研讨暨第三届网商机电子商务生态学术研讨会》论文集, 浙江大学出版社, 23-27, 2010.

6、姜元春, 合肥工业大学, 管理学院, 博士, 副研究员, 硕士生导师。负责“专题四: 大数据驱动的个性化用户建模与推荐方法”研究。

6.1 个人简介



姜元春博士近年来一直从事电子商务、商务智能、数据挖掘等方面的理论研究工作，多次担任《Decision Support Systems》、《Journal of Business Research》等学术期刊电子商务领域论文的审稿人，长期担任《管理评论》学术期刊的审稿人。先后主持国家自然科学基金青年基金、教育部人文社科基金、高等学校博士点基金（新教师类）、阿里巴巴青年学者支持计划等课题的研究工作，作为骨干成员参与了国家 973 计划项目“社交网络分析与网络信息传播的基础理论研究”子课题、国家“核高基”科技重大专项“大型网络应用及服务平台研制与示范”子课题、国家自然科学基金项目“基于网络消费者信息需求模型的网站导航问题研究”、国家“863”项目“面向制造业售后服务的商务智能关键技术的研究”的研究工作。在“Marketing Science”、“Decision Support Systems”和“IEEE Transactions on Software Engineering”等国内外学术期刊发表研究论文 30 余篇，其中 SCI 检索论文 8 篇，参与出版学术专著一部。

6.2 受教育经历（从大学本科开始，按时间倒排序）

2008/09—2011/02，美国匹兹堡大学，Katz 商学院，联合培养博士研究生

2004/09—2011/06，合肥工业大学，管理学院，博士学位

2000/09—2004/06，合肥工业大学，管理学院，学士学位

6.3 研究工作经历（按时间倒排序）

（1）工作经历

2013/12—至今，合肥工业大学，管理学院，副研究员

2009/09—2013/12，合肥工业大学，管理学院，讲师，硕士生导师

（2）主持和参加的研究课题

- [1] 在线替代品系列销售的动态协同价格策略研究，国家自然科学基金青年基金，2014.01-2016.12，主持
- [2] 在线社会性网络的信息传播及其对客户关系管理的影响，教育部人文社科青年基金，2012.01-2014.12，主持
- [3] 电子商务零售平台营销工具的组合方法研究，高等学校博士点基金（新教师类），2013.01-2015.12，主持
- [4] 社会化推荐模型与算法研究，CCF-腾讯犀牛鸟基金，2014.08-2015.08，联合主持
- [5] 基于用户行为分析的个性化推荐系统研究，阿里巴巴青年学者支持计划，2010.03-2011.03，主持
- [6] 多阶段多替代品的网络动态价格策略研究，合肥工业大学青年教师创新基金，2013.7-2014.7，主持
- [7] 基于主动协作关系的网络用户建模及模型优化方法研究，国家自然科学基金项目，2011.01-2013.12，参加
- [8] 基于网络消费者信息需求模型的网站导航问题研究，国家自然科学基金项目，



2007.01-2009.12, 参加

- [9] 新兴信息服务环境下的网络用户建模及模型优化方法研究, 高等学校博士学科点专项科研基金课题, 2010.01-2011.12, 参加
- [10] 制造业商务智能技术及产品研发-制造业商务智能支持方法研究, 国家“863”计划重点项目子课题, 2008.03-2010.09, 参加

6.4 近五年发表的与本项目相关的主要论著

(1) 期刊论文

- [1] **Yuanchun Jiang**, Jennifer Shang, Chris F. Kemerer, Yezheng Liu, Optimizing E-tailer Profits and Customer Savings: An Online Dynamic Bundle Pricing Model, Marketing Science, 30(4), 737-752, 2011.
- [2] **Yuanchun Jiang**, Jennifer Shang, Yezheng Liu, Maximizing Customer Satisfaction through an Online Recommendation System: A Novel Associative Classification Model, Decision Support Systems, 48(3), 470-479, 2010.
- [3] Xiao Liu, Yun Yang, **Yuanchun Jiang**, Preventing Temporal Violations in Scientific Workflows: Where and How, IEEE Transactions on Software Engineering, 37(6), 805-825, 2011.
- [4] **Yuanchun Jiang**, Jennifer Shang, and Yezheng Liu, Optimizing shipping-fee schedules to maximize e-tailer profits, International Journal of Production Economics, 146(2), 634-645, 2013.
- [5] **Yuanchun Jiang**, Yezheng Liu, Optimization of Online Promotion: A Profit-Maximizing Model Integrating Price Discount and Product Recommendation, International Journal of Information Technology & Decision Making, 11(5), 1-22, 2012.
- [6] **Yuanchun Jiang**, Yezheng Liu, Xiao Liu, Integrating classification capability and reliability in associative classification: A β -stronger Model, Expert Systems With Applications, 37(5), 3953-3961, 2010.
- [7] Sifeng Chen, **Yuanchun Jiang**, Yezheng Liu, Cost Constrained Mediation Model for AHP Negotiated Decision Making, Journal of Multi-Criteria Decision Analysis, 19(1-2), 3-13, 2012.

(2) 专著

- [1] 刘业政, 姜元春, 张结魁, 网络消费者行为: 理论方法及应用, 科学出版社, 2011.2.

(3) 奖励

- [1] 博士学位论文《基于智能优化方法的个性化营销策略研究》获安徽省第四届优秀博士学位论文奖, 2013 年度全国优秀博士论文提名奖。



(4) 专利

- [1] 刘业政, 姜元春, 王锦坤等, 一种基于推荐概率融合的混合推荐方法, 申请号 201310637512.4.

7、徐进, 西南交通大学, 经济管理学院, 硕士, 副教授, 硕士生导师。参与“专题一: 面向商务管理的大数据获取方法与可视化技术”研究, 重点攻克大数据的可视化问题。

7.1 个人简介

徐进副教授自任职以来以信息化理论与技术为基础, 在知识管理、智能交通等领域开展教学、研究与社会服务, 承担了数据仓库、数据挖掘、物流信息管理、公司信息战略等多门本科、研究生课程, 主持、主研国家自然科学基金、铁道部科技研究开发计划重大课题/重点课题多项, 为多个政府机构、大型国有、民营企业提供了电子政务、管理信息化、项目管理以及智能交通等方面的咨询与培训服务。研究主要关注两个方面: 信息化与项目管理, 包括项目知识管理、项目信息共享与数据挖掘、企业项目/人员情境建模、数据仓库建模等; 智能交通系统, 包括实时交通信息的收集、处理与传递, 城市交通模型建模与仿真, 智能交通系统标准等。

7.2 受教育经历(从大学本科开始, 按时间倒排序)

2000/09—2003/03, 西南交通大学, 经济管理学院, 硕士
1996/09—2000/06, 西南交通大学, 经济管理学院, 学士

7.3 研究工作经历(按时间倒排序)

(1) 工作经历

2011/07—至今, 西南交通大学, 经济管理学院, 副教授
2012/11—2013/04, 香港城市大学资讯系统系, 副研究员
2011/11—2012/02, 香港城市大学资讯系统系, 副研究员
2010/11—2011/02, 香港城市大学资讯系统系, 资深研究助理
2009/11—2010/02, 香港城市大学资讯系统系, 资深研究助理
2005/07—2011/06, 西南交通大学, 经济管理学院, 讲师
2003/04—2005/06, 西南交通大学, 经济管理学院, 助教

(2) 主持和参加的研究课题

- [1] 具备情境感知特性的项目知识推荐系统方法研究, 国家自然科学基金, 2011.01-2013.12, 主持
[2] 项目组织的能力和信任对工程伙伴关系承诺与合作影响的实证研究, 国家自然科学基金, 2010.01-2012.12, 主研



- [3] 考虑消费者选择和替代行为的多代易逝品收益管理研究，国家自然科学基金，2011.01-2013.12，主研
- [4] 京沪高速铁路建设项目质量管理体系及风险控制技术研究，铁道部科技研究开发计划重大课题，2008.01-2010.12，主研
- [5] 四川省煤炭产业集团有限责任公司管理信息化系统项目咨询服务，横向，2006.07-2009.12，主研
- [6] 四川省地方铁路局（四川铁路集团）管理信息化系统项目咨询服务，横向，2006.05-2009.10，主研

7.4 近五年发表的与本项目相关的主要论著

（1）期刊论文

- [1] Liu L, **Xu Jin**, Liao S S, et al, A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication[J], Expert Systems with Applications, 41(7), 3409-3417, 2014.
- [2] Zhang J D, **Xu Jin**, Liao S S, Aggregating and sampling methods for processing GPS data streams for traffic state estimation[J], IEEE Transactions on Intelligent Transportation Systems, 14(4), 1629-1641, 2013.
- [3] Zhang J, **Xu Jin**, Liao S S, Sampling methods for summarizing unordered vehicle-to-vehicle data streams[J], Transportation Research Part C: Emerging Technologies, 23, 56-67, 2012.
- [4] Xu K S J, Wang W, Ren J, **Xu Jin**, et al, Classifying consumer comparison opinions to uncover product strengths and weaknesses[J], International Journal of Intelligent Information Technologies (IJIIT), 7(1), 1-14, 2011.

（2）会议论文

- [1] **Xu Jin**, Zhu Jing, Stephen Shaoyi Liao, Organizational Context in Information Systems Research: Perspectives and Components. International Conference on Management and Service Science (MASS), 12-14, 2011.
- [2] **Xu Jin**, Stephen Shaoyi Liao, Organizational Context Components in Information Systems Research A Literature Review, International Conference on Engineering and Business Management (EBM), 2010.
- [3] Li S, Li J Z, Li S L, Liu D, **Xu Jin**, The overall value of Web-based hybrid models for strategic marketing decision-making: Managers' responses to the WebIntegrated and WebMarP systems[C], ICAI, 187-190, 2013.
- [4] Xu Kaiquan, **Xu Jin**, Liu Long, et al, Predict Market Share with Users' Online Activities Data: an Initial Study on Market Share and Search Index of Mobile Phone, The Pacific Asia Conference on Information Systems (PACIS), 627-633, 2010.



8、丁帅，合肥工业大学，管理学院，博士，副研究员，硕士生导师。负责“专题五：面向商务管理大数据支撑平台的核心理论与关键技术”研究。

8.1 个人简介

丁帅博士主要从事管理信息系统、云计算、可信软件等领域的研究工作。主持了国家自然科学基金青年基金项目“软件服务化背景下的云服务推荐方法研究”，并以骨干成员身份先后参与了国家自然科学基金重点项目“云计算环境下的多源信息服务系统研究”、重大研究计划项目“可信软件测试和可信性评估研究”和面上项目“云计算环境下的智能决策方法研究”的研究工作，申请国家发明专利 2 项，并在《Knowledge and Information Systems》、《Expert Systems with Applications》、《科学通报(英文版)》、《管理科学学报》、《中国管理科学》和《计算机研究与发展》等国内外期刊和学术会议上发表论文 18 篇，其中 SCI 收录 6 篇。

8.2 受教育经历（从大学本科开始，按时间倒排序）

2008/09—2011/12, 合肥工业大学, 管理学院, 博士

2005/09—2008/06, 合肥工业大学, 计算机与信息学院, 硕士

2001/09—2005/06, 安徽农业大学, 信息与计算机学院, 学士

8.3 研究工作经历（按时间倒排序）

（1）工作经历

2013/12—至今, 合肥工业大学, 管理学院, 副研究员

2012/01—2013/12, 合肥工业大学, 管理学院, 讲师

2011/08—2013/08, 美国匹兹堡大学, 计算机学院, 合作研究

（2）主持和参加的研究课题

- [1] 软件服务化背景下的云服务推荐方法研究, 国家自然科学基金, 2013.01-2015.12, 主持
- [2] 云计算环境下的多源信息服务系统研究, 国家自然科学基金重点项目, 2012.01-2016.12, 参加
- [3] 可信软件测试和可信性评估研究, 国家自然科学基金重大研究计划项目, 2008.03-2011.12, 参加
- [4] 基于社会电能信息的宏观经济决策支持系统, 江苏省电力公司委托的横向项目, 2006.09- 2008.08, 参加

8.4 近五年发表的与本项目相关的主要论著

（1）期刊论文

- [1] Shuai Ding, Shanlin Yang, et al, Combining QoS prediction and customer satisfaction



- estimation to solve cloud service trustworthiness evaluation problems[J], Knowledge-Based Systems, 56, 216-225, 2014.
- [2] **Shuai Ding**, Chengyi Xia, Kaile Zhou, et al, Decision Support for Personalized Cloud Service Selection through Multi-Attribute Trustworthiness Evaluation[J], PloS one, 9(6), e97762, 2014.
- [3] **Shuai Ding**, Xijun Ma, Shanlin Yang, A software trustworthiness evaluation model using objective weight based evidential reasoning approach, Knowledge and Information Systems, 33 (1), 171-189, 2012.
- [4] **Shuai Ding**, Shanlin Yang, Chao Fu, A novel evidential reasoning based method for software trustworthiness evaluation under the uncertain and unreliable environment, Expert Systems with Applications, 39(3), 2700-2709, 2012.
- [5] Zhou Kaile, **Shuai Ding**, Chao Fu, and Shanlin Yang, Comparison and Weighted Summation Type of Fuzzy Cluster Validity Indices, International Journal of Computers Communications & Control, 9(3), 370-378, 2014.
- [6] Juan Wang, Chengyi Xia, Yiling Wang, **Shuai Ding**, Junqing Sun, Spatial prisoner's dilemma games with increasing size of the interaction neighborhood on regular lattices, Chinese Science Bulletin, 57(7), 724-728, 2012.
- [7] 丁帅, 鲁付俊, 杨善林, 夏成遗, 一种需求驱动的软件可信性评估及演化模型, 计算机研究与发展, 48(4), 647-655, 2011.
- [8] 丁帅, 杨善林, 马溪骏, 阚红星, 面向软件演化的可信性评估自适应模型, 武汉大学学报, 35(5), 542-545, 2010.
- [9] 杨善林, 丁帅, 付超, 考虑信息源相关的软件可信性评估模型, 中国管理科学, 17(6), 163-169, 2009.
- [10] 杨善林, 丁帅, 褚伟, 一种基于效用和证据理论的可信软件评估方法, 计算机研究与发展, 46(7), 1152-1159, 2009.
- [11] 杨善林, 罗贺, 丁帅, 基于云计算的多源信息服务系统研究综述, 管理科学学报, 15(5), 83-96, 2012.
- [12] 罗贺, 杨善林, 丁帅, 云计算环境下的智能决策研究综述, 系统工程学报(录用)

(2) 发明专利(已进入实质审查阶段)

- [1] 杨善林, 丁帅, 马华伟, 胡笑旋, 罗贺, 杨露, 一种基于多属性匹配的云服务推荐方法及装置, 受理号: 201110377164.2.
- [2] 杨善林, 丁帅, 马溪骏, 余涛, 余本功, 王晓佳, 魏礼舜, 面向软件演化的可信性动态评估方法, 受理号: 201010520891.5.

9、王晓佳, 合肥工业大学, 管理学院, 博士, 讲师。参与“专题二: 面向商务管理的大数据分析与计算方法”研究, 重点攻克时序



数据和流数据分析问题。

9.1 个人简介

王晓佳博士主要从事预测、大数据挖掘等方面的研究与教学工作。在《Electrical Review》、《Journal of Applied Mathematics》、《China Communication》、《South African Journal of Industrial Engineering》、《Journal of Convergence Information Technology》、《Annals of Differential Equations》、《Control Engineering of China》、《Mathematica Applicata》、《中国管理科学》、《情报学报》、《模式识别与人工智能》、《系统仿真学报》、《应用数学》等期刊或国际会议上发表学术论文 30 余篇，其中 SCI 检索 6 篇，EI 检索 12 篇。主持与本项目相关的省级优秀青年人才基金项目 1 项，国家自然科学基金青年基金 1 项。

9.2 受教育经历（从大学本科开始，按时间倒排序）

2008/09—2012/06, 合肥工业大学, 管理学院, 博士

2004/09—2007/06, 安徽大学, 应用数学, 硕士

2000/09—2004/06, 武汉理工大学, 计算机科学与技术, 工学学士

9.3 研究工作经历（按时间倒排序）

（1）工作经历

2012/02—至今, 合肥工业大学管理学院, 讲师

2007/06—2012/02, 合肥工业大学数学学院, 讲师, 应用数学系主任

（2）主持和参加的研究课题

- [1] 不确定信息环境下基于大规模数据的趋势预测和智能决策方法研究, 国家自然科学基金, 2012.1-2014.12, 主持
- [2] 大规模数据趋势预测与智能决策方法研究, 高等学校优秀青年人才基金项目, 2012.7-2013.7, 主持
- [3] 大数据背景下面向网络舆情的政府统计决策信息供需匹配研究, 全国统计科学研究重点项目, 2014.1-2015.12, 主持
- [4] 工业与经济领域大规模数据趋势预测研究, 合肥工业大学博士学位人员专项资助基金项目, 2013.1-2014.12, 主持
- [5] 工业与经济数据的预处理机制设计与研究, 国家级创新计划项目, 2012.1-2013.12, 主持
- [6] 数据预处理的方法研究, 合肥工业大学科学研究发展基金项目, 2011.1-2012.12, 主持
- [7] 智能配用电系统海量数据预处理技术, 国家“863”重大项目, 2012.1-2014.12, 参加
- [8] 云计算环境下智能决策方法研究, 国家自然科学基金, 2011.1-2013.12, 参加
- [9] 基于社会电能信息的宏观经济决策支持系统, 企业委托课题, 2009.1-2010.12, 参加



9.4 近五年发表的与本项目相关的主要论著

(1) 期刊论文

- [1] **Xiaojia Wang**, Shanlin Yang, Wang Haijiang, etc, Dynamic GM(1,1) Model Based on Cubic Spline for Electricity Consumption Prediction in Smart Grid[J], China Communications, 7(4), 83-88, 2010.
- [2] **王晓佳**, 杨善林等, 灰色正交化方法在用电量预测中的仿真研究[J], 系统仿真学报, 22(10), 2253-2256, 2010.
- [3] **王晓佳**, 王海江, 杨善林, 基于 Markov 链的灰色正交化预测模型在智能电网用电量预测中的应用[J], 中国管理科学, 18, 46-51, 2010.
- [4] **王晓佳**, 变时滞的退化滞后型微分系统的稳定性[J], 数学的实践与认识, 40(8), 196-201, 2010.
- [5] **王晓佳**, 多时滞微分方程数值稳定性[J], 应用数学与计算数学学报, 24(1), 93-98, 2010.
- [6] **王晓佳**, 杨善林, 沈建新, 基于 Gauss 插值的正交化预测方法在智能电网用电量预测中的应用研究[J], 电力系统保护与控制, 38(21), 141-145, 2010.
- [7] **王晓佳**, 蒋威, 退化时滞中立型微分系统的特征根分布与指数稳定[J], 应用数学, 23(1), 138-144, 2010.
- [8] **Xiaojia Wang**, Shanlin Yang, Criteria On Delay-dependent Stability for Linear Descriptor System with Delay[J], Control Engineering of China, 18(2), 198-201, 2011.
- [9] **Xiaojia Wang**, The Existence, Uniqueness and Exponential Estimation of the Solution for Singular Neutral Differential Delay System[J], Mathematica Applicata, 24(2), 384-390, 2011.
- [10] **王晓佳**, 杨善林, 徐达宇, 基于改进粒子群算法的数据预测挖掘应用研究[J], 情报学报, 8(30), 840-845, 2011.
- [11] **王晓佳**, 一类复杂时滞系统的实用稳定性[J], 大学数学, 27(3), 93-97, 2011.
- [12] **王晓佳**, 变时滞区间中立型系统的稳定性新判据[J], 黑龙江大学自然科学学报, 28(1), 34-39, 2011.
- [13] **Xiaojia Wang**, Application Research on Electricity Demand Forecasting Based on Gaussian Quadrature Formula[J], Procedia Engineering, 15, 5574-5578, 2011.
- [14] **王晓佳**, 杨善林, 基于组合插值的 GM(1,1)模型预测方法的改进与应用[J], 中国管理科学, 20(2), 129-134, 2012.
- [15] **Xiaojia Wang**, On the Pointwise Degeneracy of Linear Delay Differential Equations [J], Mathematica Applicata, 25(1), 224-230, 2012.
- [16] **Xiaojia Wang**, Forecasting Modeling and Analysis of Power Engineering in China Based on Gauss-Chebyshev Formula[J], System Engineering, 5, 131-136, 2012.
- [17] **Xiaojia Wang**, Zhiqiang Chen, Shanlin Yang, Forecasting Modeling And Simulation



- Analysis of Power System in China: Based on A Class of Semi-parametric Regression Approach[J], South African Journal of Industrial Engineering, 23(2), 154-168, 2012.
- [18] 王晓佳, 张宝霆, 徐达宇, 含有压缩因子的粒子群优化灰色模型在智能电网中的应用[J], 运筹与管理, 21(3), 114-118, 2012.
- [19] Xiaojia Wang, Liqiang Hou, Zhiqiang Chen, Prediction Analysis for Short Term Load Based on GM(1,1) and TGARCH Error Correction Model in Smart Grid[J], Journal of Convergence Information Technology, 8(1), 111-117, 2013.
- [20] Xiaojia Wang, Bengong Yu, Zhiqiang Chen, Liqiang Hou, Forecasting Modeling and Analysis of Power Engineering in China Based on Semi-Parametric Regression Model[J], Electrical Review, 161-169, 2013.
- [21] 王晓佳, 杨善林, 陈志强, 大数据时代下的情报分析与挖掘技术研究[J], 情报学报, 32(6), 564-574, 2013.
- [22] Xiaojia Wang, Electricity consumption forecasting in the age of big data[J], TELKOMNIKA, 11(9), 5262-5266, 2013.
- [23] Xiaojia Wang, Zhiqiang Chen, Shanlin Yang, Electricity Consumption Prediction Based on Non-Stationary Time Series GM (1,1) Model and Its Application in Power Engineering[J], Lecture Notes in Electrical Engineering:Mechatronics and Automatic Control Systems, 237, 933-940, 2014.

10、孙见山, 合肥工业大学, 管理学院, 博士, 讲师。参与“专题三: 基于网络大数据的商业情报分析方法”研究, 重点攻克中文主题识别问题。

10.1 个人简介

孙见山博士近年来主要从事推荐系统、大数据分析和社交媒体挖掘等方面的研究工作。2014年6月和8月先后取得中国科学技术大学和香港城市大学双博士学位。先后参与了国家自然科学基金2项, 计划局软课题1项; 作为主要人员参加了深圳市互联网产业发展专项资金企业项目1项。并在《Decision Support Systems》、《Annals of Operations Research》、《Computer Journal》、《科研管理》、《科学学与科学技术管理》等国内外期刊和学术会议发表论文10余篇, 其中SCI收录3篇。

10.2 受教育经历(从大学本科开始, 按时间倒排序)

2010/09—2014/08, 香港城市大学, 商学院, 联合培养博士研究生
2009/09—2014/06, 中国科学技术大学, 管理学院, 博士学位
2005/09—2009/07, 安徽财经大学, 信息工程学院, 学士学位



10.3 研究工作经历（按时间倒排序）

（1）工作经历

2014/06—至今，合肥工业大学，管理学院，讲师

（2）主持和参加的研究课题

- [1] 精细化科研管理信息系统研究，国家自然科学基金项目，2014.01-2017.12，参加
- [2] 科研创新社区网络中基于用户特征的推荐方法研究，国家自然科学基金项目，2012.01-2015.12，参加
- [3] 社会化方法规范学科关键词，国家自然科学基金委计划局软课题，2011.07-2012.12，参加
- [4] 科研创新合作共享平台，深圳市互联网产业发展专项资金，2011.10-2013.05，参加

10.4 近五年发表的与本项目相关的主要论著

（1）期刊论文

- [1] Gang Wang, **Jianshan Sun**, Jian Ma, Kaiquan Xu, Jibao Gu, Sentiment Classification: The Contribution of Ensemble Learning, Decision Support Systems, 57, 77-93, 2014.
- [2] Yelin Fu, **Jianshan Sun**, K. K. Lai, John W. K. Leung, A Robust Optimization Solution to Bottleneck Generalized Assignment Problem under Uncertainty, Annals of Operations Research, Published online, 1-11, 2014.
- [3] **Jianshan Sun**, Jian Ma, Zhiying Liu, Yajun Miao, Leveraging Content and Connections for Scientific Article Recommendation in Social Computing Contexts, Computer Journal, Published online, 2013.
- [4] **Jianshan Sun**, Wei Xu, Jian Ma, Jiasen Sun, Leverage RAF and MapReduce to Find Experts for Research Social Network Services: A Big Data Analytics Perspective, International Journal of Production Economics, under review, 2014.
- [5] **Jianshan Sun**, Gang Wang, Jian Ma, Jiasen Sun, Mining Affective Text to Improve Social Media Item Recommendation, Information Processing & Management, 2nd round review, 2014.
- [6] Gang Wang, **Jianshan Sun**, Shanlin Yang, Catherine Larson, A Novel Enhanced Random Subspace for Sentiment Classification Based on Part-of-Speech Analysis, Information Processing & Management, 2nd round review, 2014.
- [7] Wei Xu, **Jianshan Sun**, Jian Ma, Wei Du, A Personalized Information Recommendation System for R&D Project Opportunity Finding in Big Data Contexts, Journal of Network and Computer Applications, 2nd round review, 2014.
- [8] Mingyu Zhang, Jian Ma, **Jianshan Sun**, Zhiying Liu, Supervisor Recommendation Based on Research Analytics Framework: Supplementing Indirect Relevance, Social Connection, and



Thinking Style, British Journal of Educational Technology, 2nd round review, 2014.

- [9] Xusen Cheng, **Jianshan Sun**, Yuanyuan Li, Application of a Novel Collaboration Engineering Method for Learning Design: A Case Study, British Journal of Educational Technology, Under review, 2014.
- [10] 缪亚军, 钟琪, **孙见山**, 卓越学术合作中不同身份合作效果与能力研究, 科学学与科学技术管理, 35(02), 3-10, 2014.
- [11] 缪亚军, **孙见山**, 戚巍, 卓越学术合作特征研究: 模式、强度及影响力视角, 科研管理, 已接收.

(2) 会议论文

- [1] **Jianshan Sun**, Jian Ma, Xusen Cheng, Zhiying Liu, Xiongfei Cao, Finding an Expert: A Model Recommendation System, The 34th International Conference on Information Systems (ICIS), Milan, Italy, 2013.
- [2] **Jianshan Sun**, Jian Ma, Xiaoyan Liu, Zhiying Liu, Gang Wang, Hongbing Jiang, Thushari Silva, A Novel Approach for Personalized Article Recommendation in Online Scientific Communities, The 46th Hawaii International Conference on System Sciences (HICSS), Hawaii, USA, 2013.
- [3] Chen Yang, Jian Ma, **Jianshan Sun**, Thushari Silva, Xiaoyan Liu and Zhongsheng Hua, A Weighted Topic Model Enhanced Approach for Complementary Collaborator Recommendation, The 18th Pacific Asia Conference on Information Systems (PACIS), Chengdu, China, 2014.

11、其他参与者简介。

参加本课题研究的其他学者还包括 2 名合肥工业大学“黄山学者”特聘教授 Jennifer Shang (美国匹兹堡大学运作管理专业教授) 和虞克明 (英国布鲁内尔大学统计学专业教授), 2 名西南交通大学特聘教授廖少毅 (Stephen Shaoyi Liao, 香港城市大学信息系统专业教授) 和李树良 (英国威斯敏斯特大学信息系统专业教授) 等。他们与课题团队已经建立了良好的合作关系, 并在 Marketing Science, Decision Support Systems, International Journal of Production Economics 等国际期刊上合作发表了多篇论文。虞克明教授和李树良教授将参与“专题二: 面向商务管理的大数据分析与计算方法”的研究, 并分别在抽样理论与统计检验问题、超高维问题的解决上给予指导; Jennifer Shang 教授将参与“专题四: 大数据驱动的个性化用户建模与推荐方法”的研究, 重点协助攻克个性化信息推荐的多目标优化问题; 廖少毅教授将参与“专题五: 面向商务管理大数据支撑平台的核心理论与关键技术”的研究, 为攻克支撑平台的体系结构问题提供指导。



(四) 经费申请说明

无。

(五) 其他需要说明的问题

无。

(六) 签字和盖章页中依托单位公章加盖课题申请人所属依托单位公章。合作研究单位公章盖合作研究单位的法人单位公章。

NSFC 2014



附件信息

序号	附件名称	备注	附件类型
1	企业合作意向书（百度）		其他
2	企业合作意向书（江淮汽车）		其他
3	企业合作意向书（奇瑞汽车）		其他
4	警务数据智慧应用平台鉴定意见		其他

NSFC 2014

**签字和盖章页(此页自动生成, 打印后签字盖章)**

申请人: 刘业政

依托单位: 合肥工业大学

项目名称: 面向大数据的商务分析与计算方法以及支撑平台研究

资助类别: 重大项目

亚类说明: 课题申请

附注说明: 大数据环境下的商务管理

申请人承诺:

我保证申请书内容的真实性。如果获得资助, 我将履行项目负责人职责, 严格遵守国家自然科学基金委员会的有关规定, 切实保证研究工作时间, 认真开展工作, 按时报送有关材料。若填报失实和违反规定, 本人将承担全部责任。

签字:

项目组主要成员承诺:

我保证有关申报内容的真实性。如果获得资助, 我将严格遵守国家自然科学基金委员会的有关规定, 切实保证研究工作时间, 加强合作、信息资源共享, 认真开展工作, 及时向项目负责人报送有关材料。若个人信息失实、执行项目中违反规定, 本人将承担相关责任。

编号	姓名	工作单位名称	项目分工	每年工作时间(月)	签字
1	杨善林	合肥工业大学	总体规划与研究应用方案设计	3	
2	倪志伟	合肥工业大学	大数据分析 with 计算方法	5	
3	李军	西南交通大学	大数据获取与评估	5	
4	何建民	合肥工业大学	商务分析方法	5	
5	姜元春	合肥工业大学	个性化用户建模与推荐方法	6	
6	徐进	西南交通大学	大数据可视化技术	5	
7	丁帅	合肥工业大学	大数据支撑平台	6	
8	王晓佳	合肥工业大学	大数据分析 with 计算方法	6	
9	孙见山	合肥工业大学	商业情报分析方法	6	

依托单位及合作研究单位承诺:

已按填报说明对申请人的资格和申请书内容进行了审核。申请项目如获资助, 我单位保证对研究计划实施所需要的人力、物力和工作时间等条件给予保障, 严格遵守国家自然科学基金委员会有关规定, 督促项目负责人和项目组成员以及本单位项目管理部门按照国家自然科学基金委员会的规定及时报送有关材料。

依托单位公章

日期:

合作研究单位公章1

日期:

合作研究单位公章2

日期: