



计算机科学

COMPUTER SCIENCE

基于数据融合的商务智能与分析架构研究

李爱华, 续维佳, 石勇

引用本文

李爱华, 续维佳, 石勇. 基于数据融合的商务智能与分析架构研究[J]. 计算机科学, 2022, 49(12): 185-194.

LI Ai-hua, XU Wei-jia, SHI Yong. [Framework of Business Intelligence and Analysis Based on Data Fusion](#) [J]. Computer Science, 2022, 49(12): 185-194.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[阵列光幕测试系统目标识别方法](#)

Testing System of Target Recognition Method of Array Screen

计算机科学, 2022, 49(11A): 211000109-4. <https://doi.org/10.11896/jsjcx.211000109>

[多源跨域数据融合的无线网络流量预测](#)

Traffic Prediction for Wireless Communication Networks with Multi-source and Cross-domain Data Fusion

计算机科学, 2022, 49(11A): 210800165-7. <https://doi.org/10.11896/jsjcx.210800165>

[基于多源健康感知数据动静态关系融合的疾病诊断](#)

Dynamic and Static Relationship Fusion of Multi-source Health Perception Data for Disease Diagnosis

计算机科学, 2022, 49(11A): 211100241-9. <https://doi.org/10.11896/jsjcx.211100241>

[基于规则链的网络协同制造数据融合方法研究](#)

Data Fusion Method of Network Collaborative Manufacturing Based on Rule Chain

计算机科学, 2022, 49(11A): 220300175-7. <https://doi.org/10.11896/jsjcx.220300175>

[联邦学习攻防研究综述](#)

Survey on Attacks and Defenses in Federated Learning

计算机科学, 2022, 49(7): 310-323. <https://doi.org/10.11896/jsjcx.211000079>

基于数据融合的商务智能与分析架构研究

李爱华¹ 续维佳¹ 石 勇^{2,3}

1 中央财经大学管理科学与工程学院 北京 100081

2 中国科学院大学管理学院 北京 100190

3 中国科学院大数据挖掘与知识管理重点实验室 北京 100190

(aihuali@cufe.edu.cn)

摘 要 商务智能与分析(BI&A)3.0 的出现和信息融合应用场景的拓宽增强了数据融合在商务智能研究中的重要性。越来越多经济和管理领域的研究运用了融合的思想和方法,数据融合在这些领域的应用表现出了不同于传统信息融合的特点。从信息融合和 BI&A 出发,提出了多源异构大数据背景下基于数据融合视角的 BI&A 新内涵,突出了数据融合在商务智能分析过程中的重要性。基于 WSR 系统科学方法论构建了商务智能分析“数据、信息、知识”的融合架构,使数据融合能更好地应用于经济、金融和管理等领域,为从海量多源异构数据中获取知识提供了科学依据,有利于更有效的商务智能系统的研发和实现。

关键词: 商务智能;数据融合;多源异构数据

中图法分类号 TP301;TP391

Framework of Business Intelligence and Analysis Based on Data Fusion

LI Ai-hua¹, XU Wei-jia¹ and SHI Yong^{2,3}

1 School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China

2 School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China

3 The Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China

Abstract The emergence of business intelligence and analytics(BI&A) 3.0 and the broadening application scenarios of information fusion enhance the importance of data fusion in the business intelligence. More and more researches in the fields of economy, finance and management use the idea and methods of fusion, and the application of data fusion in these fields shows characteristics different from the traditional information fusion. Considering the concepts of information fusion and BI&A, this paper puts forward the new connotation of BI&A based on the perspective of data fusion under the background of multi-source and heterogeneous big data, highlighting the importance of data fusion in BI&A. In addition, the paper constructs the fusion framework of ‘data, information and knowledge’ for BI&A based on WSR system methodology, so that the data fusion can be better applied in the fields of economy, finance and management. It provides scientific basis for acquiring knowledge from massive multi-source and heterogeneous data, and is beneficial to the development and implementation of a more effective business intelligence system.

Keywords Business intelligence, Data fusion, Multi-source and heterogeneous data

1 引言

互联网的深度发展与广泛应用使人们的生活、企业生产和政府管理彼此相连,社会、经济生活中无数参与者的活动时时刻刻都在产生大量数据,如何对其进行充分利用,挖掘其中有用的信息并将其转化成有效的知识成为了亟待解决的问题。挖掘数据价值的需求渗透到了更广泛的领域和行业。此外,海量数据跨领域、来源广、普适化、动态变化、多元异构的

新特征日益凸显,如何在传统的商务智能与分析中,将不同性质形态、标度和粒度的数据进行关联、交叉和融合成为了目前的挑战。因此,在多源异构大数据背景下,商务智能和分析有了新内涵和拓展,有必要基于数据融合对其进行更深入的研究。

数据融合又被称为信息融合,其概念起源于军事领域,它最初的含义是在军事上融合多传感器及情报信息,对目标进行定位、识别、跟踪,估计战场态势和威胁,做出正确的指挥

到稿日期:2021-11-06 返修日期:2022-04-07

基金项目:国家自然科学基金(71932008);中央财经大学新兴交叉学科建设项目

This work was supported by the National Natural Science Foundation of China(NSFC)(71932008) and Emerging Interdisciplinary Construction Project of Central University of Finance and Economics.

通信作者:石勇(yshi@ucas.ac.cn)

决策^[1]。美国国防部联合指挥实验室对其在军事领域的应用建立了 JDL 信息融合模型^[2]。随着内涵的扩大和模型的丰富,信息融合技术已被广泛应用于目标跟踪^[3]、图像处理^[4]、地质勘探^[5]、健康医疗^[6]、信息系统^[7]、通信工程^[8]等众多科学领域。数据的来源主要为传感器及机器产生的“硬数据”,一些早期的信息融合定义中表达了传感器的重要性^[9-10]。另外,一些定义包含了信息技术领域的专业术语^[11-12]。

近年来,随着应用场景的丰富,对数据融合的需求逐渐由自然科学、工业工程领域扩大到商业数据挖掘、金融风险管理的^[13]、价格预测^[14]、评价决策^[15]、政府统计等经济活动和社会管理问题,更多元、更综合的数据来源值得挖掘。大量的多源异构数据彼此割裂;数据规模与价值不匹配;数据中蕴含的实体信息及关系不断变化。此外,在经济和管理活动中,“人”作为不可缺少的参与主体,其重要性不断增强,待解决现实问题的复杂性和系统性增强,且有较多没有明确结构的“软”的因素。因此,鉴于海量数据的新特征和新领域的新需求,为了使多源数据和信息的融合更具普适性,可应用于更广泛的,尤其是新兴的经济、金融及管理等领域,需要对数据融合和商务智能分析进行再审视,以进一步丰富其内涵。已有学者结合领域知识与“信息融合”的思想,研究了经济、金融、管理等领域的应用^[13-15],但尚未有研究在理论上系统地分析数据融合在其中的新特点与研究架构。本文以信息融合、数据挖掘和商务智能为基础,提出了基于数据融合的商务智能与分析的内容和融合架构,并借鉴 WSR 系统科学方法论,结合经管领域的特点分析了架构各层次“物理、事理和人理”3 个维度的内容,强调了融合在整个商务智能分析过程中的重要作用,为数据融合在更广泛领域的应用提供了理论参考。

本文第 2 节介绍了数据融合的相关研究;第 3 节提出了基于数据融合的商务智能与分析架构;第 4 节列举了数据融合在经济、金融和管理领域的应用;最后总结全文。

2 相关工作

2.1 信息融合

当前大多数研究中,数据融合与信息融合的概念没有明显的区别^[11,16],无论是“数据”还是“信息”都包含了数值、文本、图像、音频、视频等广泛的含义。在信息融合广泛应用的领域习惯于称上述内容为“信息”,而在以数据挖掘为重要技术的商务智能与分析中,更多地用“数据”表示。商务智能与分析中的“信息”则表示从原始数据中挖掘得到的潜在内涵、模式等,通过分析和归纳有价值的“信息”可以进一步得到知识。

早期信息融合的定义指出,其主要任务是综合分析若干传感器观测到的信息^[9-10],随着研究的深入及应用的拓展,其定义也由狭窄变得宽泛。总体上,信息融合较通用的含义是:为了某一目的对来自多源的数据和信息进行组合和综合处理,以得到比单一信息更准确、可靠的估计或决策^[17]。通过信息融合,能够使多源信息优势互补、排除噪声、化解矛盾,提高信息的完整性、一致性和可信度^[11],进而增强系统的推理

认知能力,提升系统的工作决策性能。本文在 web of science 网站的 SCI 数据库中,以主题(标题、摘要、关键词)精确包含“multi-source information fusion”或者“data fusion”为条件,检索了 1990—2020 年的全部文献,绘制文章数量及研究方向数量随时间的变化情况如图 1 所示。由图 1 可知,近五年相关研究的数量增速加快,涉及的方向也更加广泛。

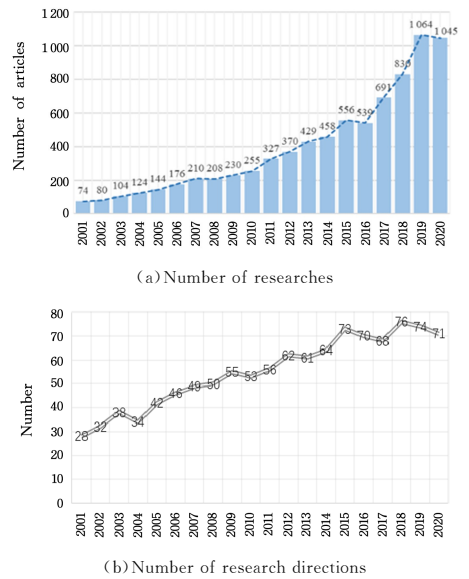


图 1 研究数量及方向

Fig. 1 Number of researches and directions

信息融合的模型主要包括结构模型^[18-19]和功能模型^[20-22]。结构模型说明了信息融合系统的工作方式,它包含 3 种基本体系:1)先将多个数据源(传感器)的数据融合,然后进行特征提取、模式判断等分析;2)先提取各数据源的特征,然后整合特征进行模式判断;3)先对各数据源提取特征并进行模式判断、身份判决后,再综合判断结果进行进一步的分析。结构模型在融合功能的部署上可以分为集中式结构、分布式结构和混合式结构^[16]。功能模型表达了信息融合系统及子系统的主要功能、作用及各部分的作用关系,典型的功能模型主要有 JDL 模型^[2]、Waterfall 模型^[20]、Omnibus 模型^[21]以及 OODA^[22]模型及其扩展等,其中应用最广泛的是 JDL、OODA 以及它们的演化模型。JDL 最初的模型主要包括第 1 级“目标评估”、第 2 级“态势评估”、第 3 级“威胁评估”和第 4 级“过程优化”4 个功能。JDL 的修正模型在“目标评估”前加入了第 0 级功能,即次目标估计,用于估计信号或特征的状态,并且将“威胁评估”调整为“影响评估”,使其应用领域从军事领域推广到民用领域。此外,随着需要解决的问题日益复杂,完全自动的信息融合系统已经难以满足实际需求,因此加入第五级认知优化功能,引入人的行为对系统进行优化^[23]。改进的 0—5 级 JDL 演化模型的框架图如图 2 所示。OODA 模型是一个由观测、定向、决策和行动 4 部分组成的闭环。针对不同的应用,OODA 产生了许多演化模型,其中扩展 OODA 可以处理可能相互影响的信息之间的融合,将用于决策的信息系统分解为多个具体的高级功能,利用 OODA 的 4 个阶段对每个高级功能进行分析评估^[24]。

的模式和知识,从而解决问题。当前,商务智能已经发展到 BI&A 3.0^[40]。BI&A 1.0 是实践中被广泛采用的技术和程序,其分析的数据大多是从公司商业关系数据库系统中收集的结构化数据。BI&A 2.0 的中心是对非结构化的文本内容的分析,互联网成为了数据的重要来源。移动设备的大规模普及使商务智能与分析进入了 BI&A 3.0,其数据主要来源于由移动设备、基于传感器的设备等收集的各种人的行为活动。当前在理论研究和实际应用中,BI&A 3.0 都仍处于新兴发展阶段。

商务智能和分析充分利用丰富的数据和许多特定领域创造的机会,产生了十分广泛的应用,主要解决经济、社会活动以及企业管理等方面的问题。例如,利用各种分析技术进行

产品个性化推荐和营销的电子商务和智能市场^[41-42];利用选择性的意见挖掘、社会网络分析等支持在线民主政治参与、透明化电子政务服务过程的智慧政府管理^[43];利用基因组大数据和支付方大数据等对患者进行电子健康记录、辅助病情监测、诊断等^[44-45];构建恐怖主义研究数据库,打击暴力、恐怖主义、网络犯罪等^[46]。

数据挖掘是商务智能分析必不可少的技术。根据 CRISP 数据挖掘参考模型指南^[47],数据挖掘过程包含理解问题、理解数据、数据准备、建模、评估和部署 6 个阶段,它们的依赖关系和主要任务如图 5 所示。由图 5 可知,仅在数据理解与准备阶段包括了数据的合成,其含义主要指将多个来源不同的结构化数据进行关联、合并和组合。

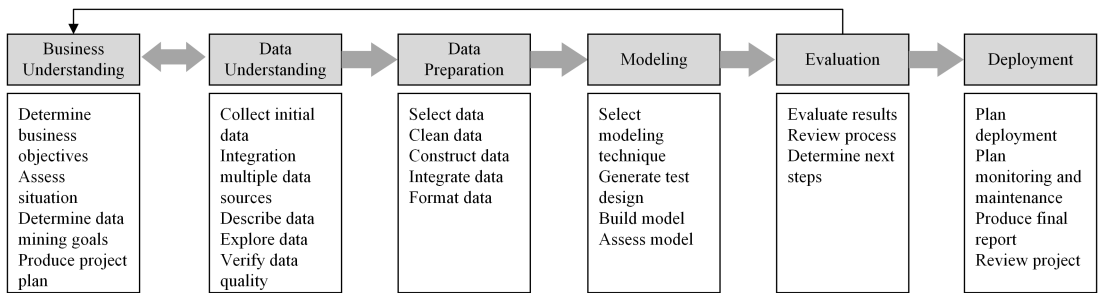


图 5 CRISP-DM 模型的流程

Fig. 5 Process of CRISP-DM

2.3 BI&A、数据挖掘和数据融合的关系分析

商务智能和分析的过程可以归纳为 3 个层次,即数据、信息和知识。数据是未经加工的,可以直接从各个存储空间获取的资源;信息是利用一定的方法和技术,对数据进行初步的统计和分析获得的潜在规律和特征;知识是信息中有价值的部分,能够为决策提供支持。综合分析传统的商务智能和信息融合的层次结构可知,前者的数据、信息和知识 3 个层次可与后者中像素级、特征级和决策级 3 个抽象层次相关联,如图 6 所示。3 个层次均由低到高,像素级融合是对原始数据的直接融合,特征级融合是对从数据中提取的特征信息进行融合,决策级融合是对数据分析得到的较为低级的判断结果进行融合。将信息融合与商务智能分析相关联,为数据融合视角下 BI&A 的“数据-信息-知识”架构的构建提供了理论基础和支持。表 2 列出了信息融合和商务智能分析的主要内容,二者的含义明显不同,信息融合强调对多样化信息的综合利用,而商务智能与分析强调从原始数据中挖掘有价值信息的过程,但是二者的 3 个抽象层次具有一定的关联。另外,信息融合和商务智能分析在数据源和应用方面也存在较显著的差异。

表 2 信息融合和商务智能与分析的比较

Table 2 Comparison of information fusion and BI&A

	Information fusion	BI&A
Level	Pixel level, feature level, decision level	Data, information, knowledge
Data source	Wireless sensors, machines and equipment, etc	RDBMS, Internet, smartphones, tablets, etc
Application	Military, communications and information engineering, automatic control, etc	E-commerce, smart market, smart government, public security, etc

利用数据挖掘技术为决策提供支持并最终形成系统,构成了完整的商务智能与分析过程。在多源异构大数据背景下,对于利用更广泛的数据来源,综合异质数据解决复杂问题的需求更加强烈,因此数据融合的地位和作用变得越来越重要,对融合的需求不仅在于数据挖掘中数据收集和准备的初始阶段,很多时候融合需要贯穿商务智能分析从数据到信息再到知识的整个过程。基于数据融合的商务智能与分析将不同层次的融合嵌入到整个 BI&A 分析过程中,使不同层次的融合成为联系密切、逐级深入的统一整体。

3 基于数据融合的商务智能与分析架构

3.1 WSR 系统科学方法论

“物理-事理-人理”系统科学方法论(简称 WSR 方法论)是对客观世界、组织和人的因素进行动态统一研究的系统方法,由系统科学家顾基发和朱志昌提出^[48],基于物质世界、系统组织和人^[49]并将其视为统一整体对系统进行研究。“物理”研究的是现实世界本身的属性和客观的规律,如获得的科学知识和构成系统的客观组成部分。研究“物理”主要是运用自然科学知识以及“硬”模型与技术工具^[50]。“事理”回答“怎么做”

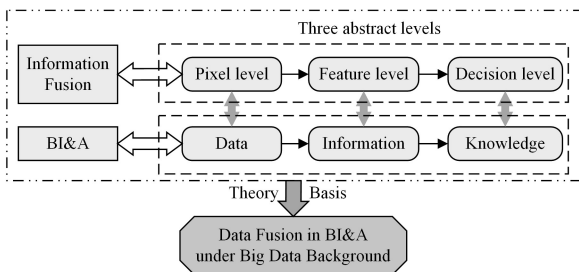


图 6 信息融合与 BI&A

Fig. 6 Information fusion and BI&A

的问题,是基于现实世界和社会的概念规律产生的干预、指导人类认识和改造世界的方法^[51],体现了人与世界的互动^[52],具有一定的主观性。“人理”是对人的研究,包括研究所处外界环境如何影响人的思想行为;研究如何发挥人的创造力和潜能,将人的理性思维的定性和阶序性及形象思维的综合性及灵活性相结合,结合已有的“物理”“事理”,实现最优的综合动态活动,以获得最大的效益和效率。人是认识和改造世界的活动主体,对事物的发展变化有很大影响,因此人与客观世界是一个整体,不应将其分隔开。“人理”涉及价值取向、文化、心理、情感、行为、目的及利益等与“人”有关的各个方面,强调了考虑问题应该将“人”的因素、人与人之间的关系纳入其中^[53]。

WSR 方法论具有综合集成的特点,在方法上(也即“事理”)是包含许多方法的方法群,而不是单一的模型和工具;在决策中,需要发挥专家群体及决策者的综合作用;在实践中,强调了协调人与人的关系,以及统筹物理、事理与人理之间的

关系^[49]。作为系统科学方法论,WSR 已经广泛应用于项目管理、交通运输、企业管理、电子商务开发、供应链管理、军事、安全等多个领域^[54-57]。

3.2 基于数据融合的“数据-信息-知识”WSR 分析

商务智能和信息融合既有相似之处,又有各自的特点,综合二者的思想,我们给出了多源异构大数据背景下基于数据融合的 BI&A 的含义^[58];基于不同领域的问题和专家经验,融合多源异构的数据并结合多种数据挖掘方法挖掘其中的信息,进一步对信息进行融合和分析,形成辅助决策的知识。数据融合视角下的商务智能与分析的核心是知识发现和辅助决策,其融合架构包含数据层融合、信息层融合和知识层融合,如图 7 所示。从原始数据资源出发,通过对数据层的融合来得到信息,在信息层融合部分对信息融合建模以得到更深层次的信息、模式、判断等,通过知识层融合对这些结果进行进一步的综合分析,以获得更高层次的决策和知识。

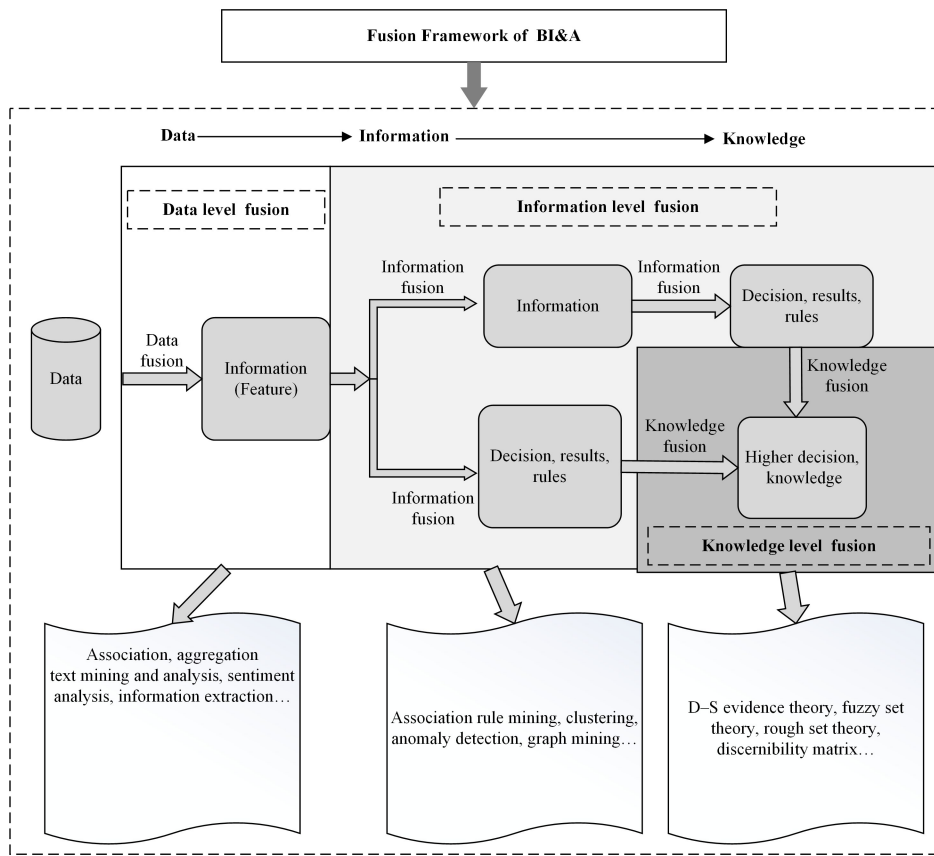


图 7 基于数据融合的 BI&A“数据-信息-知识”架构

Fig. 7 “Data-information-knowledge” framework of BI&A based on data fusion

将数据融合贯穿应用于商务智能和分析的过程使得数据融合具有了更广泛的含义,本文将其定义为:综合运用多种方法,对多源异构的原始数据、挖掘得到的内涵、模式、决策,以及其他“软因素”等进行综合、全面的处理和分析。由于应用领域的拓宽,本文中的数据融合比传统的信息融合具有更丰富的内涵。在数据源方面,传统的信息融合多为从预先设定的传感器收集数据,而数据融合的数据源更具丰富性和不确定性。数据融合不仅包括充分利用海量的、多种来源、结构各异的数据资源,还包括跨领域的模型及方法的集成和相互结合,对数据挖掘得到的信息、判断等的综合分析,以及对领域知识库和各种“软因素”等的全面考虑。图 8 给出了广义的

数据融合的内涵和外延。

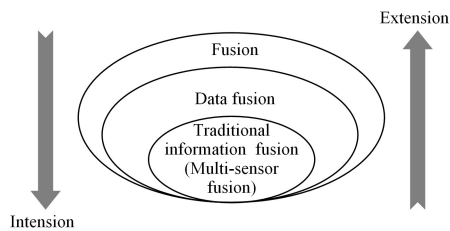


图 8 数据融合的内涵和外延

Fig. 8 Intension and extension of data fusion

基于数据融合视角的 BI&A 与传统 BI&A 最主要的

区别在于,前者强调了广义的数据融合在整个 BI&A 分析过程中的作用,使得原有商务智能问题的决策在多源异构大数据背景下更加有效。

在 BI&A 主要解决的社会、经济领域的问题中,大量数据是由人的行为活动产生的,如人的生产、消费、投资;许多主体间存在复杂密切的联系,如实体经济中的供应链、行业链、产业链,金融市场中不同市场间的风险传导。对于相关问题的研究,传统方法是在经济、管理等理论分析的基础上运用计量和统计等模型。随着互联网的普及和可获得数据的爆炸式增长,一些传统的方法难以充分利用数据资源,因此数据挖掘和人工智能等方法越来越受到研究者的青睐,“金融科技”

“智慧政府”等新概念也随之产生并受到了关注。然而,由于相关领域自身的特性,大数据技术的应用不能脱离基本的理论和原理,结果的准确性和可理解性之间需要寻求权衡,解决问题和制定决策时需要综合考虑客观状况和解决方式,以及与人有关的因素等多个方面。WSR 系统科学方法论强调了客观问题、解决方法以及“人”3 个维度的统一和协调,因此本文依据其理论,分析和拓展了基于数据融合的商务智能与分析在“数据”“信息”和“知识”3 个层次的内容,如图 9 所示,使数据融合能够更好地应用于经济、金融和管理等领域。与我们之前的研究不同^[58],此处借鉴 WSR 理论对 BI&A 的新架构进行深入分析,广义的数据融合则作为架构中的关键技术。

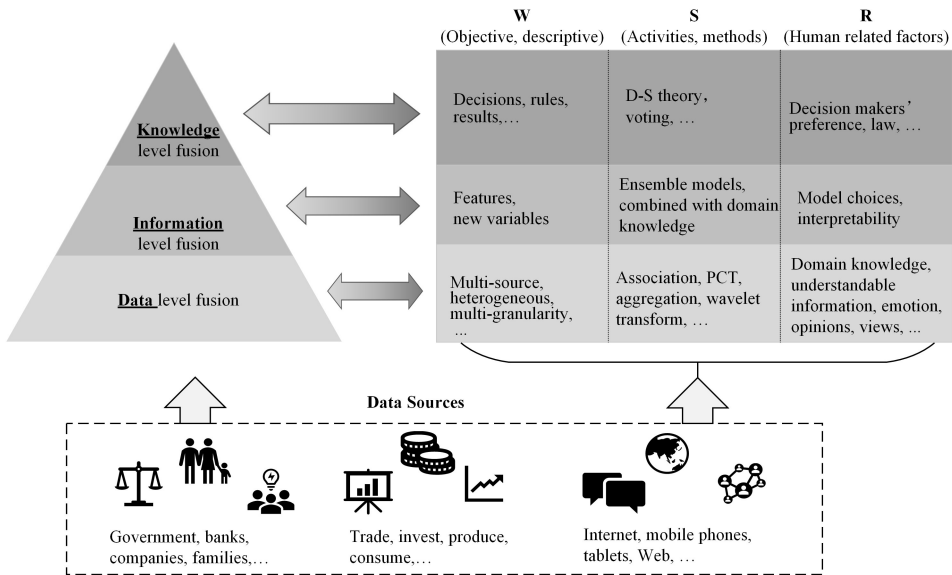


图 9 融合的 3 个层次

Fig. 9 Three levels of fusion

商务智能与分析的数据层融合的主要任务是多来源数据的收集和信息的提取。数据的理解和准备是数据挖掘过程中重要的初始步骤,然而对于应收集的数据,相关研究都未给出详细的说明。在广泛应用信息融合的目标识别、追踪及控制等问题中,数据来源主要为预先设定的无线传感器,具有比较清晰的限定。然而,经济、金融和管理等领域往往涉及社会系统,实体较多且关系复杂,很多问题中存在较长的因果链条。因此,在跨媒体的大量数据来源中筛选数据具有较大的不确定性,对于同样的问题,不同的研究者可能会选择不同的数据。例如研究银行的系统性风险,既可以从微观数据入手,选择多家银行分析其资产负债结构等,也可以将中观数据作为研究对象,直接选取银行业风险度量指标。有研究指出,数据的选取既要根据实际问题,又要结合领域知识和专家经验^[59]。借鉴 WSR 理论的思想,数据的获取和选择可以从“物”“事”“人”3 个维度进行考虑,即与问题或事物直接相关的客观数据、各种行为活动产生的数据以及与“人”的因素相关的数据。例如,在对某股价进行预测时,可以选取表示该公司客观情况的财务数据,如市值、净利润、股票流通股数等;股票交易产生的数据,如历史价格、成交量等;以及投资者的评价和预期信息,如社交媒体的股评、新闻信息等。在进行产品个性化推荐时,可以选择产品本身的相关信息(价格、质量检测

等)、购买产生的信息(销售量等)以及消费者的评价等。

异质数据的转化和信息的提取是 BI&A 数据层融合的另一重要任务。在传统信息融合的像素级融合中,相同介质传感器的数据可以直接融合。然而在很多 BI&A 的问题中,不同来源的数据形式不同,例如不同表中的结构化数据,不同粒度的时间序列数据,以及政策文件、媒体新闻、网络舆论等非结构化文本、图像数据等。对于形态各异的数据,很难直接对它们进行建模和分析,因此需要将不同形态的数据通过组合、聚合、关联、文本挖掘等技术转化为相同形式的、能合并分析的信息。异质数据转化和信息提取与信息融合中的“特征提取”不完全相同,特征提取的目的主要在于学习原始数据中的特征,而 BI&A 数据层融合中信息提取的目的在于处理不同来源、形态、粒度的数据,为信息层的融合和模型构建提供输入,它不仅是表征学习,更强调所提取信息的可理解性和实际意义。此外,信息提取还包括根据领域知识,综合利用原始数据构建新的综合性的变量和指标,以反映更全面、更深入的信息,对信息的提取需要充分考虑其实际含义和解释。例如,利用公司原始财务数据构建反映盈利能力、偿债能力、成长能力等的新指标;通过对文本数据的挖掘和分析,构建反映文本可读性、一致性、真实性等具有现实意义的表征向量。虽然有时深度挖掘深层次特征能够提高结果的准确性,但

同时也丧失了特征的可解释性,降低了结果的说服力。

商务智能与分析的信息层融合是将数据层融合得到的信息通过构建模型进行合并分析。信息层融合的“物理”是上一阶段数据层融合得到的综合性的信息,“事理”是融合模型的构建,可能用到的技术有关联规则挖掘、分类聚类算法、机器学习、异常检测算法等,以及多种模型的集成与组合,如模糊神经网络、遗传神经网络以及各种集成算法等。经济、管理等领域模型的构建不能脱离领域知识和基本原理,如何将数据挖掘及信息融合的技术与传统的计量、统计分析等方法有机结合,仍然是新兴的研究问题和难点。信息层融合的“人理”在于人们对模型的选择和理解,信息的含义和模型的底层逻辑不可忽视,因此在解决实际问题时需要提高有关主体的参与度,构建领域知识库,增强模型的可解释性和可理解性。在社会和经济活动的问题中,结果的准确性并非唯一追求的目标,对其原因的分析以及与现实情况的联系对领域专家而言十分重要。

商务智能与分析的知识层融合是最高层次的融合,是对信息层融合得到的有价值的模式、判断等结果,综合专家意见等更多因素进行再融合,以得到更高层次的知识,为决策提供支持。知识层融合是将低层次阶段融合得到的笼统的知识转化为可领悟的知识,通过对表层知识的推理、归纳得到显示或隐式的深度关系和知识,面向需求和决策提供知识服务^[60]。知识层融合的灵活度和容错性高,抗干扰能力强,常用方法有结合 D-S 证据推理的专家意见综合评估、投票系统、区分矩阵、产生式规则,以及管理学中的方法等。由于最终的决策和知识的认知都离不开“人”,如专家意见能够为最终决策提供有效参考,决策者偏好可能对决策造成影响,不同社会文化环境和法律背景等可能导致决策产生偏差等,因此在知识层融合中需要综合考虑与“人”有关的多方面因素,将其与所获分析结果相结合,为决策提供更有力的支持。例如,在金融欺诈检测中,首先可以通过数据融合、人工智能的方式进行初步筛查和判断,而最后的罚没情况需要进一步结合法律法规、被告人申诉内容、情节轻重等进行判决。与数据层融合和模型层融合相比,知识层融合的层次更高,适用范围更广,能够通过挖掘初级层面的知识来得到更深层次的内涵,其对多决策结果的综合也能进一步提高结果的准确性和稳定性。

总体上,基于数据融合的商务智能与分析中,数据层、信息层和知识层的融合层次越来越高。在“数据-信息-知识”框架中每个融合层次都需要综合考虑“物理”“事理”和“人理”3个方面,各层次中3方面因素有其各自的含义与内容。“数据-信息-知识”商务智能和分析的3个融合层次与传统信息融合抽象得到的数据级、特征级和决策级3个融合级别主要有以下不同:首先,商务智能与分析中3个融合层次的含义和适用性更广,特别是对于解决经济、金融、管理等领域的问题,各层次的融合相比传统信息融合增添了新的内涵和特征,特别是与“人理”有关的新内容。另外,基于数据融合的商务智能与分析强调数据层融合、信息层融合和知识层融合的连续性、统一性和实践中的递进性。相比之下,信息融合的3个层次相对独立,有些问题的解决仅运用数据级(像素级)融合,将原始数据融合之后提取特征并构建模型;有些则仅运用特征

级融合,先对数据提取特征然后建立模型。然而,由于商务智能与分析所解决的社会活动中的问题较为复杂和系统,综合运用3个融合层次,先对形态各异的数据进行融合,再对特征构建集成模型,最后将专家经验、领域知识等与结果相结合,使融合贯穿分析的全过程,以更好地提供决策支持。

4 基于数据融合的 BI&A 的应用

近年来,越来越多的学者将数据融合应用于解决经济、金融、政府及企业管理领域的问题。虽然很多文献并未明确提及“数据融合”或“信息融合”的概念,但研究中或综合利用了多种来源、不同形态的数据,或采用了多种方法构建集成模型,都体现了“数据融合”的思想。

在经济领域,学者们收集了价格、销量等商品市场数据以及特定研究问题的相关数据,综合数据挖掘技术与经济模型,研究了产量预测^[14]、人口预测^[61]、经济效益评价^[62]等问题。Padilla等^[14]融合地区市场销售数据、气象数据等空间分布的多源数据网络,提取不同市场产品间的关联知识来改进空间预测技术,为农业生产监测提供了一个数据融合系统,以实现科学生产和更准确的销售预测。Mao等^[62]基于数据融合,提出了海运港口货物集散运输管理的经济效益评价模型,将经济统计信息和海运港口货物集散运输管理中的关联规则特征相结合,利用模糊子空间调度构建了经济效益评价的控制模型,具有较高的评价能力和较强的可信度。Zuo等^[63]结合已声明的偏好数据和隐含的偏好数据来估计澳大利亚各种类型的水权销售的价格弹性,通过不同偏好数据的比较为数据融合方法提供了强有力的证据。越来越多的研究表明,数据融合给经济分析和相关研究提供了新的思路和视角^[61-63]。

在金融领域,Figini等^[13]在纵向和生存期模型中进行数据融合,综合了量化的资产负债率和从非结构化文本中挖掘的定性信息并将其按权重组合,得到中小企业违约概率,其实验结果显示,综合多源异构数据并且利用多种融合模型能够得到更优的违约评估。此外,资本市场是金融市场的重要组成部分,一些学者利用融合的思想对资本市场的价格预测^[64]、风险管理^[65]等问题进行了研究。Lee等^[64]运用多模态深度学习模型,综合国际股票市场间的转移信息对股票价格进行预测,结果显示融合模型的预测效果远优于单一模型。此外,Wang等^[65]在股价操纵检测中运用了融合的思想,在数据层综合了高频交易的时间序列数据与公司财务状况数据,在信息层上采用了集成模型,在决策上综合了时间序列数据检测结果与财务数据检测结果,实验结果显示,综合交易数据与财务数据的集成模型的检测效果优于仅用单一信息的单一模型。Kong^[66]收集了收盘价、最高价等股票量价数据、所属板块信息以及相关新闻文本信息,利用栈式自编码、LINE 提取量价信息的深层次特征向量,利用 Bi-LSTM 等模型对文本进行情感分析,将不同信息的特征向量进行拼接以得到融合信息向量,并结合神经网络和注意力机制构建模型预测股价,结果显示融合量价信息与新闻文本信息在股市预测任务上取得了显著的效果。

在管理领域,学者们利用数据融合的思想对决策管理、商业价值分析等问题进行了大量的研究^[67-71]。Franceschini

等^[15]研究了多智能体偏好排序融合形成共识排序的问题,从多个角度检验共识排序与输入数据的一致性,为各种环境下的决策提供支持。Hou 等^[67]在利用商务智能技术构建矿业集团运营决策系统方案中指出,可以采用联邦数据库系统、数据库、数据访问中间件等数据融合技术,充分利用原有系统历史数据,实现异构系统间的信息集成,从而为信息平台搭建提供理论支撑和模型方法。Chen 等^[68]综合多元异构数据对失信被执行人进行特征提取,利用曲线排齐算法融合两种数据特征,有效提高了特征信息数据的提取量,保证了后续失信被执行人特征识别的准确性。Ji 等^[69]融合用户社交关系和评论文本信息,从中提取主题信息并将用户划分为不同的社区,利用机器学习方法建立模型得到用户对商家的评分,根据评分结果对商家排序并将排序后的结果推荐给用户,实现了更准确的个性化推荐。Huang 等^[70]融合多源数据,通过企业竞争对手画像指标体系构建、数据采集、数据融合与分析、画像构建等环节,提出了基于多源数据融合的企业竞争对手画像构建模式,为解决海量竞争对手与有限竞争情报资源之间的矛盾提供了参考。Ding^[71]结合酒店客户需求大数据融合调度和特征挖掘模型以及融合特征检测方法,对相关信息特征进行提取和挖掘,提高了酒店对客户需求的预测能力。

由于来源广泛、可获得性增强的海量数据提供了条件和资源,且传统的方法对分析大量、高维、结构各异的数据存在局限性,因此将数据融合方法与技术应用于经济、金融、管理领域的相关实际场景已成为重要趋势。然而,尚未有完整、系统的理论对传统的信息融合与其在新兴应用领域中产生的新内容进行分析。通过对经济、金融和管理领域具有数据融合思想的应用研究进行总结和抽象,本文利用 WSR 系统方法分析了基于数据融合视角的广义商务智能与分析,为应用研究提供了理论补充和参考。

结束语 随着 BI&A 3.0 的出现和信息融合应用领域的拓宽,数据融合在数据挖掘和商务智能中的重要性与日俱增。越来越多针对经济、金融和管理领域问题的研究运用了融合的思想和方法,数据融合在其中的应用表现出了区别于传统信息融合应用领域的特点。本文综合信息融合和商务智能分析的概念,总结和抽象了数据融合在经济管理等领域的应用研究,提出了多源异构大数据背景下基于数据融合视角的 BI&A“数据-信息-知识”的融合架构,突出了广义数据融合在整个智能分析和问题解决过程中的重要作用。BI&A 融合架构包括 3 个融合层次:数据层、信息层和知识层。由于社会生产活动缺少清晰的结构和流程,具有较强的综合性和灵活性,因此借鉴 WSR 的理论,对 3 个融合层次中“物、事、人”因素进行了深入的分析。

数据层融合可以从 3 个维度考虑数据的选取,将不同形态的数据转化成能共同建模的形式,从原始数据中提取可理解的信息或构造新的特征。信息层的融合在于多种方法的集成和综合、经济学原理等领域知识和新兴技术的结合,以及准确性与可解释性之间的平衡。BI&A 知识层的融合是最高层次的融合,除了对初步结果或判定的融合,还包括考虑偏好、经验等与“人”有关的因素,挖掘更深度的显示和隐式知识,为决策提供支持。

未来的研究中,各融合层次领域知识库的构建、新兴技术的因果推断以及高层次融合中如何融合更多“人”的因素等,是值得进一步深入思考的问题。

参考文献

- [1] WALTZ E, LLINAS J. Multi sensor data fusion[M]. London: Artech House Radar Library, 1990.
- [2] WHITE J F E. A model for data fusion[C]//Proceedings of the 1st National Symposium on Sensor Fusion. 1988:149-158.
- [3] DING Y, YU X, ZHANG J, et al. Application of linear predictive coding and data fusion process for target tracking by Doppler Through-Wall Radar [J]. IEEE Transactions on Microwave Theory and Techniques, 2018, 67(3): 1244-1254.
- [4] SUSPERREGI L, ARRUTI A, JAUREGI E, et al. Fusing multiple image transformations and a thermal sensor with kinect to improve person detection ability[J]. Engineering Applications of Artificial Intelligence, 2013, 26(8): 1980-1991.
- [5] SHEN H, WU J, CHENG Q, et al. A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(3): 862-874.
- [6] LI K F. Smart home technology for telemedicine and emergency management[J]. Journal of Ambient Intelligence and Humanized Computing, 2013, 4(5): 535-546.
- [7] ATAT R, LIU L, WU J, et al. Big data meet cyber-physical systems: A panoramic survey [J]. IEEE Access, 2018, 6: 73603-73636.
- [8] NALLAGONDA S. Data fusion-aided cognitive radio network over generalised fading channels[J]. Electronics Letters, 2019, 55(5): 285-287.
- [9] WALD L. Some terms of reference in data fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(3): 1190-1193.
- [10] BLACKMAN S, POPOLI R. Design and analysis of modern tracking systems[M]. London: Artech House Publishers, 1999.
- [11] CHEN K W, ZHANG Z P, LONG J. Multisource information fusion: key issues, research progress and new trends[J]. Computer Science, 2013, 40(8): 6-13.
- [12] DONG Z R. Information fusion[J]. Command Control & Simulation, 2001, 23(7): 27-36.
- [13] FIGINI S, GIUDICI P. Statistical merging of rating models[J]. Journal of the Operational Research Society, 2011, 62(6): 1067-1074.
- [14] PADILLA W R, GARCÍA J, MOLINA J M. Knowledge extraction and improved data fusion for sales prediction in local agricultural markets [J/OL]. Sensors, 2019, 19(2): 286. <https://doi.org/10.3390/s19020286>.
- [15] FRANCESCHINI F, MAISANO D. Checking the consistency of the solution in ordinal semi-democratic decision-making problems[J/OL]. Omega, 2015, 57: 188-195. <https://doi.org/10.1016/j.omega.2015.04.014>.
- [16] PENG D L, WEN C L, XUE A K. Theory and application of

- multi-sensor and multi-source information Fusion[M]. Xi'an: Xidian University, 2010.
- [17] ZHAO J, CUI Z S, XU J M, et al. The essence and core technology of information fusion[J]. *Command Control & Simulation*, 2003, 25(8): 38-42.
- [18] ZHAO Z G, LI J L, WANG K. The concept, structure and efficiency of battlefield situation assessment[J]. *Journal of China Academy of Electronics and Information Technology*, 2010, 5(3): 226-230.
- [19] DE VIN L J, HOLM M, NG A. The information fusion JDL-U model as a reference model for virtual manufacturing[J]. *Robotics and Computer-Integrated Manufacturing*, 2010, 26(6): 629-638.
- [20] HARRIS C J, BAILEY A, DODD T J. Multi-Sensor data fusion in defence and aerospace [J]. *Aeronautical Journal*, 1998, 102(1015): 229-244.
- [21] BEDWORTH M, O'BRIEN J. The Omnibus model: a new model of data fusion? [J]. *Aerospace & Electronic Systems Magazine IEEE*, 2000, 15(4): 30-36.
- [22] BLASCH E P, BRETON R, VALIN P, et al. User information fusion decision making analysis with the C-OODA model[C]// 14th International Conference on Information Fusion. New York: IEEE Press, 2011: 1-8.
- [23] LIGGINS M, HALL D, LLINAS J. *Handbook of multisensor data fusion: theory and practice (Second Edition)* [M]. New York: CRC Press, 2008.
- [24] SHAHBAZIAN E, BLODGETT D, LABBÉ P. The extended OODA model for data fusion systems[C]// The 4th International Conference on Information Fusion. 2001.
- [25] WANG R, JI W, LIU M, et al. Review on mining data from multiple data sources[J/OL]. *Pattern Recognition Letters*, 2018, 109: 120-128. <https://doi.org/10.1016/j.patrec.2018.01.013>.
- [26] ESCAMILLA-AMBROSIO P J, MORT N. A hybrid Kalman filter-fuzzy logic architecture for multisensor data fusion[C]// Proceeding of the 2001 IEEE International Symposium on Intelligent Control (ISIC'01) (Cat. No. 01CH37206). New York: IEEE Press, 2001: 364-369.
- [27] DU H, LV F, LI S, et al. Study of fault diagnosis method based on data fusion technology[J/OL]. *Procedia Engineering*, 2012, 29: 2590-2594. <https://doi.org/10.1016/j.proeng.2012.01.356>
- [28] RÓVID A, REMELI V. Towards raw sensor fusion in 3D object detection[C]// IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI). New York: IEEE Press, 2019: 293-298.
- [29] LIANG M, YANG B, WANG S, et al. Deep continuous fusion for multi-sensor 3d object detection[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich: Springer, 2018: 641-656.
- [30] KUZNETSOVA P, ORDONEZ V, BERG T L, et al. TRE-TALK: composition and compression of trees for image descriptions[J]. *Transactions of the Association for Computational Linguistics*, 2014, 2(9): 351-362.
- [31] DU Q, XU H, MA Y, et al. Fusing infrared and visible images of different resolutions via total variation model[J/OL]. *Sensors*, 2018, 18(11): 3827. <https://doi.org/10.3390/s18113827>.
- [32] CORONA I, GIACINTO G, MAZZARIELLO C, et al. Information fusion for computer security: State of the art and open issues[J]. *Information Fusion*, 2009, 10(4): 274-284.
- [33] LIU P F, ZHANG P L, ZHANG J, et al. Subject oriented Web information fusion model[J]. *Library and Information Service*, 2011, 55(8): 40-43.
- [34] ALVES S F R, ROSARIO J M, FERASOLI F H, et al. Conceptual bases of robot navigation modeling, control and applications in Robot Navigation[M]. IntechOpen Access Publisher, 2011.
- [35] LATHUILIÈRE S, MASSÉ B, MESEJO P, et al. Deep reinforcement learning for audio-visual gaze control [C] // 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE Press, 2018: 1555-1562.
- [36] DEEN M J. Information and communications technologies for elderly ubiquitous healthcare in a smart home [J]. *Personal Ubiquitous Comput*, 2015, 19(3/4): 573-599.
- [37] JAMES A P, DASARATHY B V. Medical image fusion: A survey of the state of the art[J/OL]. *Information Fusion*, 2014, 19: 4-19. <https://doi.org/10.1016/j.inffus.2013.12.002>.
- [38] LIANG X, HU P, ZHANG L, et al. MCFNet: Multi-layer concatenation fusion network for medical images fusion[J]. *IEEE Sensors Journal*, 2019, 19(16): 7107-7119.
- [39] HANNAN B, ZHANG X, SETHARES K. IHANDs: Intelligent health advising and decision-support agent [C] // 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). New York: IEEE Press, 2014, 3: 294-301.
- [40] CHEN H, CHIANG R H L, STOREY V C. Business intelligence and analytics: from big data to big impact[J]. *MIS Quarterly*, 2012, 36(12): 1165-1188.
- [41] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734-749.
- [42] JOHNSON J, NG Y. Enhancing long tail item recommendations using tripartite graphs and Markov process[C]// Proceedings of the International Conference on Web Intelligence. New York: ACM, 2017: 761-768.
- [43] CHEN H. AI, E-government, and Politics 2.0[J]. *IEEE Intelligent Systems*, 2009, 24(5): 64-86.
- [44] HANAUER D A, ZHENG K, RAMAKRISHNAN N, et al. Opportunities and challenges in association and episode discovery from electronic health records[J]. *IEEE Intelligent Systems*, 2011, 26(5): 83-87.
- [45] MAGBOO M S A, CORONEL A D. Data mining electronic health records to support evidence-based clinical decisions[J/OL]. *Innovation in Medicine and Healthcare Systems, and Multimedia*, 2019: 223-232. <https://doi.org/10.1007/978-981-13-8566-722>.
- [46] BRANTINGHAM P L. Computational criminology[C]// 2011 European Intelligence and Security Informatics Conference. New York: IEEE Press, 2011.

- [47] CHAPMAN P, CLINTON J, KERBER R, et al. CRISP-DM 1.0 step-by-step data mining guide[M]. Chicago:SPSS Inc, 1999.
- [48] GU J F, ZHU Z C. Knowing Wuli, sensing Shili, caring for Renli; methodology of the WSR approach[J]. Systemic Practice and Action Research, 2000, 13(1): 11-20.
- [49] GU J F, TANG X J, ZHU Z X. Review of Wuli-Shili-Renli system methodology[J]. Journal of Transportation Systems Engineering and Information Technology, 2007, 7(6): 51-60.
- [50] ZHU Z C. Enlightenment of international exchange on Wuli-Shili-Renli methodology [C] // Systems Engineering, Systems Science and Complexity Research Proceeding of 11th Annual Conference of Systems Engineering Society of China. Research Information Ltd. ,2000, 149-164.
- [51] GU J F, GAO F. Wuli-Shili-Renli system methodology based on the perspective of management science[J]. Systems Engineering-Theory & Practice, 1998, 18(8): 2-6.
- [52] KOU X D, GU J F. A twenty-five-year review of WSR methodology: origin, connotation, comparison and outlook [J]. Management Review, 2021, 33(5): 3-14.
- [53] GU J F, TANG X J. From ancient system thoughts to modern oriental systems methodology[J]. Systems Engineering-Theory & Practice, 2000 20(1): 90-93.
- [54] GU J F, TANG X J. Designing a water resources management decision support system: an application of the WSR approach [J]. Systemic Practice and Action Research, 2000, 13(1): 59-70.
- [55] SHE L Z. An empirical analysis of WSR system for large-scale engineering projects[J]. China Civil Engineering Journal, 2006, 39(6): 111-114.
- [56] ZHANG Q, XUE H F. An analytical model for environmental safety based on WSR methodology [J]. China Soft Science, 2010, 25(1): 165-174.
- [57] YE W, LIU X Y. Empirical study on factors affecting the system of science and technology innovation talent development under the perspective of WSR[J]. Science and Technology Management Research, 2017, 37(9): 36-43.
- [58] LI A H, XU W J, SHI Y. A New data fusion framework of business intelligence and analytics in economy, finance and management[C] // 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). New York: IEEE Press, 2020: 940-945.
- [59] PENG Y, KOU G. Research on theoretical framework of data mining based on domain knowledge[C] // The third (2008) Chinese Management Annual Conference-Information Management Conference Proceedings. Changsha: Chinese Research Council of Modern Management, 2008: 1242-1250.
- [60] MENG X F, DU Z J. Research on the big data fusion: issues and challenges[J]. Journal of Computer Research and Development, 2016, 53(2): 231-246.
- [61] JEVTIC P, REGIS L. A continuous-time stochastic model for the mortality surface of multiple populations[J/OL]. Insurance: Mathematics and Economics, 2019, 88: 181-195. <https://doi.org/10.1016/j.insmatheco.2019.07.001>.
- [62] MAO Y, GAN S. Economic evaluation model of freight distribution management in maritime port [J]. Journal of Coastal Research, 2019 (Special Issue No. 93): 1059-1065.
- [63] ZUO A, ANN W S, ADAMOWICZ W L, et al. Measuring price elasticities of demand and supply of water entitlements based on stated and revealed preference data[J]. American Journal of Agricultural Economics, 2016, 98(1): 314-332.
- [64] LEE S I, YOO S J. Multimodal deep learning for finance: integrating and forecasting international stock markets[J]. The Journal of Supercomputing, 2020, 76(10): 8294-8312.
- [65] WANG Q, XU W, HUANG X, et al. Enhancing intraday stock price manipulation detection by leveraging recurrent neural networks with ensemble learning[J]. Neurocomputing, 2019, 347: 46-58.
- [66] KONG D S. Research on technologies of stock market prediction based on quantity-price and sentiment analysis [D]. Harbin: Harbin Institute of Technology, 2019.
- [67] HOU J, HU N L, LI G Q, et al. Solution of operation and decision-making system oriented to mining group based on business intelligence[J]. Computer Integrated Manufacturing Systems, 2016, 22(1): 202-212.
- [68] CHEN G, YING Y H, WANG Y. Research on feature extraction of persons subject to enforcement for trust-breaking based on multivariate heterogeneous data fusion[J]. Legality Vision, 2020 (32): 187-188.
- [69] JI Z Y, PI H Y, YAO W N. A hybrid recommendation model based on fusion of multi-source heterogeneous data[J]. Journal of Beijing University of Posts and Telecommunications, 2019, 42(1): 126-132.
- [70] HUANG X B, ZHANG M X. Construction of enterprise competitor portrait based on multi-source data[J]. Journal of Modern Information, 2020, 40(11): 13-21.
- [71] DING D. Research on the application of big data in hotel customer demand forecast under tourism background[J]. Journal of Qiqihar University (Natural Science Edition), 2020, 36(6): 90-94.



LI Ai-hua, born in 1978, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include big data management and decision making, and financial risk management.



SHI Yong, born in 1956, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include data mining and knowledge management.