

大数据环境下多源信息融合的理论与应用探讨^{*}

■ 化柏林 李广建

北京大学信息管理系 北京 100871

摘要: [目的/意义]从理论、方法、技术与应用的视角研究大数据环境下的多源信息融合问题。[方法/过程]探索多源信息融合的理论基础与科学依据,梳理多源信息融合的问题与现象,对多源信息融合的技术与方法进行归纳总结,对构建大数据环境下的多源信息融合应用进行探讨。[结果/结论]大数据环境下的多源信息融合理论包括相关性原理、多元表示原理、意义建构理论等,多源信息融合方法包括统一标识、数据比对、异构加权等一系列过程以及多种分析建模方法,多源信息融合在国家层面、城市或行业组织层面、企业机构层面都有着广泛的需求与应用。

关键词: 大数据 多源信息融 情报分析

分类号: G250

DOI: 10.13266/j.issn.0252-3116.2015.16.001

1 引言

多源异构是大数据的基本特征之一,多源数据融合也成为大数据分析处理的关键环节,多源数据融合成为大数据领域重要的话题与研究方向。中国计算机学会(China Computer Federation, CCF)大数据专家委员会秘书长程学旗在 2014 年大数据大会上发布了《中国大数据技术与产业发展白皮书(2014 年)》,对 2015 年大数据发展趋势进行预测,总结为:融合、跨界、基础、突破。从中可以看出,融合成为最为显著的发展趋势^[1]。“互联网+”等概念的提出进一步表明,信息融合、产业融合、经济融合等多维度融合是时代发展的要求,是顺应大数据社会发展的必然。通过“互联网+”的互动融合,可实现行业的模式转变与效率提升。

多源信息融合在大数据时代具有非常重要的价值与意义。通过多源信息融合,有利于进一步挖掘数据的价值,提升信息分析的作用;通过多源信息交叉印证,可以减少信息错误与疏漏,防止决策失误。可以说,在大数据时代,融合成为一个重要的理念、一个广泛渗透于各行业的现象、一种涉及多学科的新常态,需要深入研究。

研究大数据环境下的多源信息融合问题需要从理论、方法、技术与应用 4 个视角展开探讨。在理论层面探讨多元表示原理、相关性原理、意义构建等支撑理论,在方法层次涉及到贝叶斯、D-S 证据、神经网络等算法,在技术层面既涉及线上线下数据融合、传感数据与社会数据融合等问题,也涉及唯一识别、异构加权等技术细节,通过理论指导与方法技术的支撑,实现企业级、城市区域级以及国家级多源信息融合的应用,见图 1。

2 多源信息融合的理论

新的渠道、新的载体不断产生新的数据类型,如何应对复杂多变的多源数据,成为大数据环境下一个重要的问题。要想解决这个问题,就需要剖析信息融合的定义与内涵,深入研究多源信息融合的机制与机理,探寻多源信息融合的本质与规律,从而解释多源信息融合的现象,指导多源信息融合的实践。

信息融合最早应用于军事领域,后来在传感器、地理空间、情报分析等多个领域得到了应用与发展。关于信息融合主要有以下几种定义:①信息融合是一种多层次、多方面的数据处理过程,对来自多个信息源的

^{*} 本文系中国博士后科学基金特别资助项目“大数据环境下的多源信息融合机理与方法研究”(项目编号:2015T80027)和国家社会科学基金重点项目“大数据环境下的计算型情报分析方法与技术研究”(项目编号:14ATQ005)研究成果之一。

作者简介: 化柏林(ORCID:0000-0001-9248-6455),博士后, E-mail: huabolin@pku.edu.cn; 李广建(ORCID:0000-0002-2897-6246),教授,博士生导师。

收稿日期:2015-07-13 修回日期:2015-08-04 本文起止页码:5-10 本文责任编辑:刘远颖

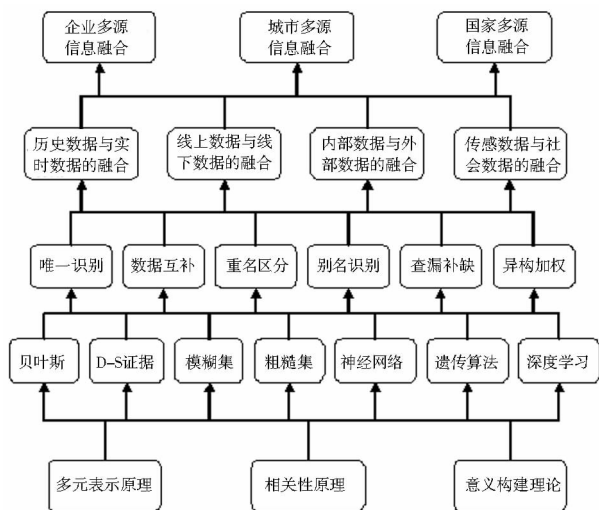


图 1 多源信息融合的理论方法体系

数据进行自动检测、关联、相关、估计及组合等处理。

②信息融合是研究利用各种有效方法把不同来源、不同时间点的信息自动或半自动地转换成一种能为人类或自动的决策提供有效支持的表示形式^[2]。

③信息融合是处理探测、互联、估计以及组合多源信息和数据的多层次多方面过程,以便获得准确的状态和身份估计、完整而及时的战场态势和威胁估计^[3]。

④多源数据融合是指由不同的用户、不同的来源渠道产生的,具有多种不同的呈现形式(如数值型、文本型、图形图像、音频视频格式),描述同一主题的数据并为了共同的任务或目标融合到一起的过程^[4]。这些研究都反映了多源信息融合的含义或理念,但并没有形成较为系统的信息融合理论。

对多源信息融合虽然已有不少研究,但多源信息融合的理论仍比较缺乏。在这种背景下,探寻与借鉴已有的理论,对多源融合的现象进行解释,就成为一项有重要意义的研究。本文从逻辑语义、语法结构以及形式表示 3 个方面展开讨论,在这个层面的理论基础主要包括相关性原理、多元表示原理以及意义建构理论等,见图 2。不同的理论对多源信息融合有不同的支撑,多元表示更容易揭示多源信息的外部特征,是表现形式;相关性原理可以反映事物或事物要素及属性之间的关联关系,是结构与关系;意义建构在认知与语义的层面,反映内在逻辑。

2.1 多元表示原理

不同的人对同一数据会使用不同的数据表现形式,例如论文题目在中国知网中用“篇名”,而在万方数据中用“题名”,还有的数据库用“标题”,用多元表示原理可以解释同一数据的多种表述。多元表示是信

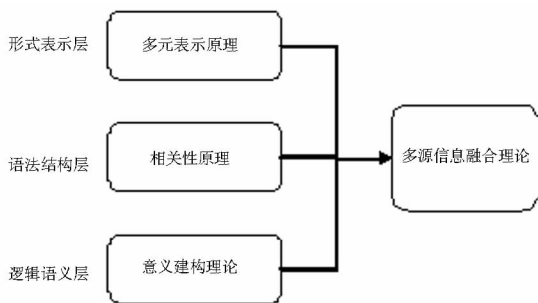


图 2 多源信息融合的支撑理论框架

息活动中的一种普遍现象,存在于信息活动的各个环节和各个阶段,这在互联网环境下更加明显,最为典型的的就是大众标签,同样的内容,不同的人标记的标签就不一样。多元表示包括两种形式:一种是来自于不同认知行动者解读的在认知上不同的表示;一种是来自于同一个认知行动者的在功能上不同的表示。从认知意义上来讲,即便是同一组认知行动者(如标引员)在面对同一信息对象或情景时,其成员之间也可能会表现出认识上的不一致性或解释上的多样性^[5]。在众筹众创时代,多元表示原理可以很好地揭示多源异构数据的多种表象,并寻找共同的语义内容与关联。

2.2 相关性原理

大数据分析比传统的数据分析更加注重相关性分析,这一方面由于大数据具有数据规模体量大、多源异构等特点,进行简单直接的相关性分析比复杂的因果分析有更高的计算效率;另一方面,通过对大数据的相关性分析,可以直接发现一些有用的关联关系,如购物篮中的同被购买商品,发现这样的现象足以提高经济效益,没必要非得弄清楚同被购买的原因。用相关性可以解决多源信息融合中的一些问题,包括主题相关、要素相关、任务相关、情境相关^[6]等。多源信息融合需要根据相关性原理判定数据之间的相关关系与关联程度以及数据源与任务情境之间的相关性。相关性作为大数据与情报学的共性原理之一^[7],在大数据环境下的情报研究中有着重要的应用价值。

2.3 意义建构理论

布伦达·德尔文于 1972 年提出以用户为中心的意义建构理论(sense-making theory),认为知识是主观、由个人建构而成,而信息寻求是一种主观建构的活动,在线检索的过程是一连串互动、解决问题的过程,是一种解释沟通、信息与意义之间关系的概念性工具。无论是认知层面的内在行为还是以过程为主体的外显行为,都允许个体在空间和时间上设计或建造自己的行为^[8]。意义建构的行为是种沟通行为,而信息恰恰就

是人与人之间的最有效沟通的载体形式。信息的产生、组织、加工、标引、检索、传播与利用等过程都涉及用户的参与,而这些用户参与的行为与表现可以用意义建构来解释。根据皮亚杰的理论,人在与环境相互作用的过程中,不断建构和修正原有的知识结构,不同的知识结构所决定的信息形式与内容就会不同,而意义建构理论有助于揭示不同数据源对任务目标的支撑作用。

3 多源信息融合的方法与技术

有关多源数据融合的方法、技术等,学界已有一些研究。B. Khaleghia 等人对多源数据融合进行了全面的综述,包括数据融合的概念、价值、难点以及现有的方法^[9]。R. R. Yager 使用投票的方法解决数据之间的冲突,提出一种多源数据融合的框架^[10]。F. Naumann 等人认为需要通过“模式匹配、重复侦测、数据融合”3个步骤来解决多源数据的不一致性及其数据冲突问题^[11];陈科文等人则从数据处理、系统设计、融合模型、融合方法等方面梳理当前多源信息融合的关键问题与应用进展^[12]。K. Chopra 建立了一个多源信息融合的软件工具模型,并对情报分析的不确定性评价进行了探讨^[13]。A. Marco 等人用重组认知集成方法实现智能应用的高层次融合,通过定义一个由原语、功能和模型构建的涉及语义、时态和地理空间等多维信息的集成框架,在现有的信息融合模型之间实现桥接,提出了一套实现统一的高层次融合智能应用程序的方法,通过案例研究演示了在知识发现和预测精度改进方面的应用^[14]。

多源数据融合涉及到很多具体的方法与技术,但这些技术方法都是零散的、针对某个具体问题或应用场景的,不足以应对当前对多源、异构、跨界信息融合的需求。为了满足多源信息融合的全面需求,就需要对这些方法与工具进行分析与试用,研究各种技术方法的适用性、优缺点以及相互之间的关系等,从而形成多源信息融合的技术方法体系。

3.1 多源信息融合的表现形式

多源信息融合有多种来源与表现形式。在相关性、多元表示等原理的支撑下,多源信息融合有哪些形式与表征也是重要的研究内容。多源信息在进行融合的过程中首先要转化为机器可读的数据,从数据的角度进行大规模的融合,其表现形式包括传感数据与社会数据的融合、历史数据与实时数据的融合、线上数据与线下数据的融合、内部数据与外部数据的融合。

3.1.1 内部数据与外部数据的融合 从数据来源方面来讲,数据分为企业内部数据与外部数据。内部数据一般是高质量的、与业务逻辑紧密联系的;外部数据又包括可免费获取的(如互联网数据)以及购买或合作的数据。企业人事管理系统、财务管理系统、进销存管理系统、ERP、CRM 等系统里存有大量的高质量数据,是多年业务运行累积的结果。除了自身拥有的数据以外,在大数据环境下,还需要整合一些互联网数据,如微博数据、社交数据、历史交易数据,新闻、股票论坛、公司公告、行业研究报告、行情数据等。通过整合行业和机构内部的各种数据源,并积极借助互联网等外部数据,实现内外数据的多源融合。

3.1.2 历史数据与实时数据融合 从时效性来看,经过多年的信息化,组织机构或企业已积累了相当数量的数据,而新运行的系统与网络又不断产生新的数据,通过新数据可以监测实时状态,而纵观历史数据可以发现规律,从而实现对未来的预测。仅有实时数据无法探其规律,仅有历史数据,也无法知其最新状态,要想更好地发挥数据价值,既要重视历史数据的累积与利用,又要不断获取鲜活的新数据。数据表示的是过去,但表达的是未来^[15],只有把历史数据与实时数据融合起来,才能通过历史看未来。

3.1.3 线上数据与线下数据融合 随着企业的互联网化,越来越多的传统企业也开始互联网化,在原有的线下数据基础上又产生了大量的线上数据,通过线上数据获取实时状态、客户信息,通过线下的活动进一步推广完善,实现线上数据与线下数据的有效对接。过去传统行业的领域知识是靠行业内不断摸爬滚打积累起来的,而互联网化之后的行业领域知识将是海量的用户行为数据中分析和挖掘出来的。通过互联网产业促进线上线下融合发展,为经济转型发展提供新途径和新模式。

3.1.4 传感数据与社会数据融合 传感器、射频、监控器以及其他通信设备每天产生大量的数据,这些“硬数据”以物理信号的形式传到云端服务器上,一般具有良好定义的特征;社交网络、经济活动运行、政府信息管理又产生大量的社会数据,称之为“软数据”,是人为生成的或知识性的数据(如语言文字或图形符号),这些数据带有很大的不确定性和模糊性。物理信号数据反映机器设备运行的状态,社会数据很好地反映了人们在社会运行中的各项活动。把这些信号数据与社会数据融合到一起,就可以更好地揭示自然规律与人类的各项活动。

3.2 多源信息融合的流程与技术

多源信息融合的实现包括数据级(或信号级、像素级)融合、特征级融合和决策级融合等 3 个层次,这 3 个层次的融合分别是对原始数据、从中提取的特征信息和经过进一步评估或推理得到的局部决策信息进行融合。数据级和特征级融合属于低层次融合,而高层次的决策级融合涉及到态势认识与评估、影响评估、融合过程优化,等等。H. V. Jagadish 教授等人认为,在数据时代,很多人只注重“分析”和“建模”,而忽略了其他步骤的重要性^[16],如数据的清洗与融合。

融合的过程中有些共性的流程,也存在一些差异化的过程。针对每一步过程,有多种解决问题的方法,不同的方法又有着不同的技术实现。因此有必要通过梳理多源信息融合的流程,总结多源信息融合的方法并集成多源信息融合技术,形成多源信息融合的技术方法体系。多源信息融合主要涉及数据唯一识别、数据记录滤重、字段映射与互补、重名区分、别名识别、异构数据加权等多个方面,每个方面都涉及具体的技术细节与处理方法。例如,数据归一就包括全称与缩写、同义词的转换、缩略语与全称的转化、机构的改名与合并、公司的重组与兼并等^[17]。这就需要深入研究方法之间的逻辑关系,对方法的上下位类、同族、替代、改进等关系进行归纳总结,形成方法体系;对各种技术工具进行比对与试用,分析技术工具之间的共性与优缺点,探讨技术工具的集成与应用,从而形成一套多源信息融合的技术方法体系。

3.3 多源信息的融合的算法与模型

在多源信息融合之后,需要进行深入的分析挖掘。对多源数据进行融合挖掘,就需要建立各种数据挖掘模型,包括用户聚类分析、消费模式挖掘、行业标杆对比、预警分析、客户路径分析等,以用户和业务为核心,对用户的相关维度进行数据挖掘,构建用户和业务的属性与特征库,服务于各种业务需求。

多源信息融合的算法包括简单方法、基于概率论的方法、基于模糊推理的方法以及人工智能算法等。简单算法如等值融合法、加权平均法等。基于概率论的信息融合方法,如贝叶斯方法^[18],D-S 证据理论^[19]等,其中贝叶斯方法又包括贝叶斯估计、贝叶斯滤波^[20]和贝叶斯推理网络等。D-S 证据理论是对概率论的推广,既可处理数据的不确定性,也能应对数据的多义性。基于模糊逻辑的信息融合方法,如模糊集、粗糙集^[21]等方法,这些方法在处理数据的模糊性、不完全性和不同粒度等方面具有一定的适应性和优势。混合方法包括模糊 D-S 证据

理论、模糊粗糙集理论等,可以处理具有混合特性的数据。人工智能计算方法,如神经网络^[22]、遗传算法、蚁群算法、深度学习算法^[23]等,可以处理不完善的数据,在处理数据的过程中不断学习与归纳,把不完善的数据融合为统一的完善的数据^[12]。

4 多源信息融合的典型应用

多源信息融合在传感器领域应用比较广泛,但该领域主要考虑信号数据的获取与融合,对诸如互联网等社会数据的融合关注不够,这为大数据环境下的多源信息融合带来了一定的局限性。另一方面,情报领域的专家学者对大数据给情报领域带来的机遇与挑战进行了广泛探讨,并认为融合是大数据环境下情报领域要着力解决的问题之一。李广建等人认为大数据时代下的情报研究应从单一领域情报研究转向全领域情报研究、综合利用多种数据源、注重新型信息资源的分析、强调情报研究的严谨性和情报研究的智能化 5 个方面^[24]。张志强认为,科技文献情报机构需要构建适应数据密集型科学范式需求的新型范式,着力开展专业型、计算型、战略型、政策型和方法型等“五型融合”的科技情报研究新范式^[25]。郑彦宁等人认为“多源信息”与“多元方法”既可保证产业竞争情报分析的全面性,也可通过信息源及不同方法分析结果间的交叉验证,还可进一步提高情报产品的有效性^[26]。无论从学科理论研究,还是从情报实践研究的角度,关注多源信息融合已成为情报的一个重要特点与趋势。

4.1 多源信息融合的国家级应用

从宏观层面来看,新的时代特征给国家的战略布局、城市的运行管理以及企业的运营管理都带来了新的挑战与机遇。从工业化与信息化的两化融合,到云大移(云计算+大数据+移动互联网)的融合,再到今天备受关注的“互联网+”,无不体现着融合的理念、要素、技术在当今时代的重要性。不仅如此,国家更需要通过提升信息化水平,促进经济结构转型升级,从源头上与过程中有效防止腐败,通过全面信息化支撑全面的深化改革,而全面信息化就要打破藩篱,建立统一的数据中心。这就需要对多来源、跨区域、跨部门的数据进行融合。例如,2014 年年初备受关注的“单独两孩”政策在各地逐步落地,而这项政策的出台,就是相关机构和部门通过“国家人口宏观管理与决策信息系统(PADIS)”,整合来自公安、民政、卫生、财税、教育、劳动与社会保障等各个部门的各类信息,通过对人口增长与社会承载发展进行长时间的跟踪与模拟,为适

度的政策制定选择一个合适的时间窗口,从而进行辅助决策。

4.2 多源信息融合的城市区域级应用

从中观层面来看,城市综合治理涉及到很多数据,如城市部件数据、人口数据、房屋数据及各种业务管理数据,这些数据从不同视角反映人物、事件或活动的相关信息。不同的系统产生不同的数据,不同的部门提供不同来源的数据,如何将这些多源异构数据汇聚到一起,是城市综合治理中信息管理的一个重要问题。通过收集和汇聚各类政务信息,建立大数据决策分析模型,对海量数据进行统计分析、时间序列、监测分析、场景预测、模式发现、趋势预测等,一方面可以增强对城市管理与规划、重大突发事件、自然灾害及重要舆情的监测、预警、研判和处置能力,提高城市综合治理与社会服务能力;另一方面,有利于优化产业结构与布局,提升区域创新能力与竞争力。

4.3 多源信息融合的企业级应用

从微观层面来看,现代企业运营需要全面地了解市场动态、技术前沿、竞争格局与对手进展,需要在充分整合企业内外数据、结构化与非结构化数据、线上数据与线下数据的前提下,充分利用这些不同来源、不同类型、不同结构的全信息,深入分析竞争对手的核心竞争力与弱点所在,全面扫描当前的竞争环境与态势,为企业的科学运营与决策提供有利的基础与保障,只有这样,才能在激烈的市场竞争中立于不败之地。例如,通过把搜索引擎的检索日志、电商网站的交易记录以及社交网站的人际网络与活动动态等多源信息融合汇聚到一起,可以较为全面地认识并掌握某个用户信息行为特征,从而准确地刻画出“千人千面”,更全面地揭示用户特点,根据用户特点与应用场景构建更好的数据挖掘模型,实现数据价值的最大化。

从国家战略到城市管理再到企业运营,都需要多源信息的支撑,在运用云计算、大数据、移动互联网等技术时,需要着力解决多源、异构信息在跨界中的融合与应用问题。同一个事实或规律可以隐藏在不同数据源中,不同数据源也可以揭示同一个事实或规律的不同侧面,既为分析结论的交叉印证提供了契机,也使得分析者有意识地融合汇集各种类型的数据。通过多渠道信息的汇聚,减少信息错误疏漏,提高科学运营与决策能力,提高市场竞争与环境适应能力。

5 结语

多源数据融合在情报领域有着广泛的应用,是大

数据时代的突出特点,对多源数据融合的基础理论、技术方法以及实践应用进行研究具有重要意义,从多源信息融合的整体研究与发展来看,其具有以下特点及发展趋势:①多源信息融合与大数据一样,是一个实践驱动领域,这导致了从数据层次或实践角度探讨数据融合技术或方法的研究较多,但随着融合现象与融合实践的增加,未来从理论层面对多源信息融合的机理与机制的研究将会进一步丰富。②多源信息融合起源于传感器等物理信号数据的融合,但随着社会交流形式的丰富,社会数据成为一种非常重要的数据来源,把物理数据与社会数据融合到一起进行多维分析与展示将成为必然。③情报领域向来注重多源信息融合,利用多源信息进行交叉印证与关联分析,在大数据环境下针对海量的多源异构信息,构建一个基于多源信息融合的情报工作模式与方法体系将是情报领域一个新的发展趋向。

本文的研究还是初步的,未来笔者将深入地探讨多源信息融合的理论机制,构建一套成熟完备的多源信息融合方法技术体系,并选取合适的领域进行实证研究,以更好地适应数据环境与现实问题的需求。

参考文献:

[1] 李国杰,程学旗. 大数据研究:未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊,2012(6):647-657.

[2] Boström H, Andler S F, Brohede M, et al. On the definition of information fusion as a field of research[R]. School of Humanities and Informatics, University of Skövde, 2007.

[3] 李广建,化柏林. 我们的大数据时代[M]. 北京:中国人事出版社,2015.

[4] 唐卫平,颜冰. 多传感器信息融合技术在网络雷阵中的应用[J]. 水雷战与舰船防护,2005(2):25-29.

[5] 张新民. 多元表示与情报学[J]. 情报理论与实践,2009,32(7):23-28.

[6] 张新民,罗卫东. 相关性情报学[J]. 情报理论与实践,2008,31(1):12-14,64.

[7] 马费成. 论情报学的基本原理及理论体系构建[J]. 情报学报,2007,26(1):3-13.

[8] 曾丹. 基于意义建构信息利用偏差弥合的释义[J]. 武汉理工大学学报(社会科学版),2012,25(4):636-641.

[9] Khaleghi B, Khamis A, Karray F O. Multisensor data fusion: A review of the state-of-the-art[J]. Information Fusion,2013,14(1):28-44.

[10] Yager R R. A framework for multi-source data fusion[J]. Information Sciences, 2004,163(1):175-200.

[11] Naumann F, Bilke A, Bleiholder J, et al. Data fusion in three steps: Resolving inconsistencies at schema-, tuple-, and value-level

- el[J]. IEEE Data Engineering Bulletin, 2006, 29(2):21-31.
- [12] 陈科文,张祖平,龙军. 多源信息融合关键问题、研究进展与新动向[J]. 计算机科学,2013, 40(8):6-13.
- [13] Chopra K. Information fusion for intelligence analysis[C/OL]. Proceedings of the 38th Hawaii International Conference on System Sciences,Hawaii,2005. [2015-08-26]. <https://www.computer.org/csdl/proceedings/hicss/2005/2268/04/22680111a.pdf>.
- [14] Solano M A, Ekwaro-Osire S,Tanik M M. High-level fusion for intelligence applications using recombinant cognition synthesis[J]. Information Fusion,2012, 13(1): 79-98
- [15] 涂子沛. 数据之巅[M]. 北京:中信出版社,2014.
- [16] Jagadish H V, Gehrke J, Labrinidis A, et al. Big data and its technical challenges[J]. Communications of the ACM, 2014,57(7): 86-94.
- [17] 化柏林. 多源信息融合方法研究[J]. 情报理论与实践,2013(11):16-19.
- [18] 张金槐. 多源信息的 Bayes 融合精度鉴定方法[J]. 国防科技大学学报,2001(3):93-97.
- [19] Si Lei, Wang Zhongbin, Tan Chao, et al. A novel approach for coal seam terrain prediction through information fusion of improved D-S evidence theory and neural network[J]. Measurement,2014,54(8):140-151.
- [20] Garcia E, Hausotte T, Amthor A. Bayes filter for dynamic coordinate measurements - Accuracy improvement, data fusion and measurement uncertainty evaluation[J]. Measurement,2013, 46(9): 3737-3744.
- [21] Lin Guoping, Liang Jiye, Qian Yuhua. An information fusion approach by combining multigranulation rough sets and evidence theory[J]. Information Sciences, 2015, 314: 184-199.
- [22] Safari S, Shabani F, Simon D. Multirate multisensor data fusion for linear systems using Kalman filters and a neural network[J]. Aerospace Science and Technology,2014, 39(12): 465-471.
- [23] Suk Heung-Il, Lee Seong-Whan, Shen Dinggang. Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis[J]. Neuroimage, 2014,101: 569-582.
- [24] 李广建,杨林. 大数据视角下的情报研究与情报研究技术[J]. 图书与情报,2012(6):1-8.
- [25] 张志强. 论科技情报研究新范式[J]. 情报学报,2012,31(8): 788-797.
- [26] 郑彦宁,刘志辉,赵筱媛,等. 基于多源信息与多元方法的产业竞争情报分析范式[J]. 情报学报,2013,32(3):228-234.

作者贡献说明:

化柏林:撰写与修改论文;

李广建:指导论文选题与写作,并提出修改建议。

Discussion on Theory and Application of Multi-Source Information Fusion in Big Data Environment

Hua Bolin Li Guangjian

Department of Information Management, Peking University, Beijing 100871

Abstract: [Purpose/significance] To study the problem of multi-source information fusion in big data environment from the perspective of theory, method, technology and application. [Method/process] Explores the theory foundation and scientific basis of multi-source information fusion, and clarifies the problems and phenomenon of the multi-source information fusion, then summarizes technologies and methods of multi-source information fusion, to present applications of multi-source information fusion in big data environment. [Result/conclusion] Theories of multi-source information fusion include relevance principle, polyrepresentation principle, and sense-making theory. Methods of multi-source information fusion include a series of processes, such as unified resource, data comparison, heterogeneous weighted and a variety of analysis and modeling methods. At last, widely application and demand of multi-source information fusion are presented at the national level, city or level of industry, enterprise organization level.

Keywords: big data multi-source information fusion intelligence analysis

下 期 要 目

□国外高校图书馆数据馆员服务模式研究

(蒋丽丽 陈幼华 陈琛)

□ORCID 在机构知识库建设中的应用研究

(王海花 黄金霞 刘雪梅)

□美国大学图书馆版权服务调查及启示

(叶秀明)

□科学数据共享中的邻避现象及应对研究

(周姗姗 翁苏湘 毕强等)