

· 研究论文 ·

大数据时代知识融合的支撑理论架构^{*}

于梦月 申 静

(北京大学信息管理系,北京,100871)

[摘 要] 设计大数据时代知识融合的支撑理论架构可以有效指导知识融合的实现路径和服务模式。基于知识融合的三阶段论,通过梳理和剖析知识融合前、融合过程和融合后的相关支撑理论,设计大数据时代知识融合的支撑理论架构。该架构为大数据时代的知识融合全过程提供了相关理论支撑,由三个层次、三个阶段、十五个理论组成,包括知识融合前融合的原因与融合对象的选取,知识融合过程中知识的转换、表示、计算与融合算法的实施,知识融合后知识的可视化与效果评价等相关理论。

[关键词] 大数据时代 知识融合 融合全过程 三阶段论 理论架构

[中图分类号] G251 [文献标识码] A [文章编号] 2095-2171(2020)03-0018-09

DOI: 10.13365/j.jirm.2020.03.018

Supporting Theory Framework for Knowledge Fusion in the Era of Big Data

Yu Mengyue Shen Jing

(Department of Information Management, Peking University, Beijing, 100871)

[Abstract] The supporting theory of knowledge fusion in the era of big data can effectively guide the implementing paths and service patterns of knowledge fusion. Based on the three-stage theory of knowledge fusion, the supporting theoretical framework of knowledge fusion in the era of big data is designed by combing and analyzing the relevant support theory before, during and after the fusion of knowledge. This architecture provides relevant theoretical support for the whole process of knowledge fusion in the era of big data. It consists of three levels, three stages, and fifteen theories, including various related theories such as the reasons for fusion and the selection of fusion objects before knowledge fusion, the transformation, representation, calculation and fusion of knowledge in the process, the visualization and effect evaluation of knowledge after knowledge fusion.

[Keywords] Big data era; Knowledge fusion; Fusion process; Three-stage theory; Theoretical framework

^{*} 本文系“2019年全国情报学博士生学术论坛”推荐录用论文。

[基金项目] 本文系2015年度国家社会科学基金重大项目“大数据时代知识融合的体系架构、实现模式及实证研究”(15ZDB129)的研究成果之一。

[作者简介] 于梦月,博士研究生,研究方向为知识管理;申静(通讯作者),教授,博士生导师,研究方向为知识管理与服务创新,Email:jshen@pku.edu.cn。

本文引用格式:于梦月,申静.大数据时代知识融合的支撑理论架构[J].信息资源管理学报,2020,10(3):18-26.

1 引言

知识融合是从多个异构的网络资源中定位和抽取知识并对其进行转换的过程,并将这种知识组合结果应用于问题求解^[1]。知识融合是面向知识服务和决策问题,以多源异构数据为基础,在本体库和规则库的支持下,通过知识抽取和转换获得隐藏在数据资源中的知识因子及其关联关系,进而在语义层次上组合、推理、创造出新知识的过程^[2]。传统的知识融合往往是少量的、静态的知识,在固定的学科领域或应用场景下,以人为应用主体的知识汇聚与融合过程。大数据时代,海量的数据中蕴含着庞杂的知识,这些知识规模巨大、类型众多、价值密度低且产生速度快。为了高效地获取和利用这些知识,以及为某一领域的问题求解构造更有效的知识资源,大数据时代的知识融合往往是多种来源渠道、结构与表现形式各异的、动态迭代更新并实时有效的,同时在广泛的学科领域或应用场景下,以计算机和人交互使用为目的,融合后的知识具有统一的知识模式,可以更完整、更确定、更好地满足不同层次用户的需求。

近年来,有不少学者就知识融合的技术问题进行了较多研究,但对知识融合的支撑理论进行的系统研究不多。本文通过梳理和剖析知识融合的相关支撑理论,来设计大数据时代知识融合的支撑理论架构。

2 相关研究

知识融合是继数据融合、信息融合后的高阶融合模式,是大数据时代满足用户需求和提升知识服务水平与服务质量的有效途径。因此,大数据时代的知识融合成为国内外学者关注的焦点。

国外对知识融合的研究最早可追溯到 20 世纪 80 年代后期,如 Douglas Lenat 开始的 Cyc 项目^[3]。关于知识融合的研究文献发文量最多的国家是美国,研究方向集中在计算机科学、工程与电信等领域,研究内容主要关注在特定领域的融合过程^[4-5]、融合体系结构的组建^[6-7]和融合结果的优化^[8]等。关于知识融合的理论研究,Andrade 等运用 Demster-Shafer 理论和不精确的 Dirichlet 模型进行知识融合的规则研究^[9];Suzuki 等基于模糊集理论和

D-S 理论的不确定性知识进行数据融合的方法研究^[10];Mahmoud 等运用模糊集理论应对决策支持系统中的不确定性和不精确性进行多源信息融合的方法研究^[11];Niu 等基于粗糙集和粒度计算的知识获取机制集成为多源知识融合的思维进化算法^[12];Xu 等基于信息扩散理论提出了融合-知识测度的知识融合评价系统^[13];Preece 等将本体作为通用知识交换格式的约束,建立了用于支持分布式、多异构的知识融合和转换的 KRAFT 架构^[14];Kuo 等提出了基于本体的知识融合框架^[15];Dong 等指出知识可以表示为三元组形式,并可以被提取、融合形成知识库^[16];Bryl 等认为通过在数据源之间设置 RDF 链接,可以简化 web 数据的集成,解决数据源之间的冲突,为进一步知识的融合与推理打下基础^[17]。

国内学者在 2000 年左右开始关注知识融合问题,发文量呈阶梯状增长趋势。近 5 年研究集中在教育、计算机、图书与情报、管理学科领域,研究内容主要包括知识融合概念探析^[2,18]、体系架构研究^[19-20]、应用研究^[21-23]等。关于知识融合的理论研究,化柏林等从逻辑语义、语法结构以及形式表示三个方面,认为相关性原理、多元表示原理以及意义建构理论对多源信息融合起到了理论支撑的作用^[24];张心源等认为目前国内外较为成熟的知识融合规则有 D-S 证据理论、贝叶斯网络、本体论、模糊集理论等^[25];王敬东等通过本体论的知识表示,设计了基于知识协同的知识处理模型^[26];刘晓娟等认为知识融合的核心特征是发挥多来源数据的协同效果,利用本体可将融合来源进行语义细化^[2];韩翔宇等基于本体的知识组织方法,提出实现多领域的知识融合方法^[27];蒋黎黎等基于粒度计算理论对本体模块进行分解与重构,所构建的领域知识本体更加简洁和语义一致^[28];王曰芬等提出面向知识生态重构的知识融合支撑理论与方法,包括用户认知与行为理论,大数据相关理论与方法,知识组织、分析与服务理论与方法,语义、本体等相关理论与方法,可视化与人机交互理论与方法等^[20];唐旭丽等综合考虑环境情境、个人情境和领域本体三个方面,提出了一种情境本体驱动的多源知识融合框架^[29];宋艳辉等提出的基于知识空间理论的新型知识融合方

法,将知识空间中的每个知识对象用相应的本体对象和元知识集表示,并把知识对象之间的距离由结构和语义关系来量化,以及基于信息扩散理论的评价机制可以提高融合知识的准确度^[30]。

综上,国外对于知识融合的研究起步早、成果较多,国内虽处于起步发展阶段,但研究热情高涨。目前,国内外关于知识融合的理论研究偏重于探索如何运用相关理论进行知识融合的方法,尚未关注到如何从知识融合的全过程探讨相关理论的支撑作用。

3 研究设计

(1) 研究过程与研究方法

基于知识融合的三阶段论^[25],根据大数据时代的知识融合特点,采用内容分析法剖析系统调研的知识融合相关理论,运用归纳法将这些相关理论划分到知识融合之前、知识融合过程、知识融合之后三个阶段,并阐释这些理论对知识融合各个阶段的支撑作用。在此基础上,设计大数据时代知识融合的支撑理论架构。

(2) 数据采集与处理分析

首先,以 CNKI 和万方数据库作为中文文献的主要来源,检索时间为 2019 年 6 月 16 日,分别以“知识融合”“信息融合”“数据融合”为主题词进行检索。在 CNKI 中,相应的中文文献分别有 1359 篇、13425 篇和 21464 篇,在万方数据库中,文献类型为期刊论文、学位论文和会议论文,经匹配后得到的中文文献分别有 1498 篇、19101 篇和 19490 篇。去除重复文献,选取以“知识融合”为主题词的相关文献 2388 篇进行人工判读,梳理知识融合的相关理论。其次,以 WOS(web of science)作为英文文献的主要来源,检索时间为 2019 年 6 月 16 日,以“知识融合”为主题词进行检索得到文献 10689 篇,其中高被引论文有 63 篇,经人工判读后发现这些文献与知识融合主题的相关性不大。然后又以“知识融合”为标题进行检索得到文献 246 篇,经人工判读后发现这些文献与知识融合主题的相关性较高,因此选择后一方案进行梳理分析。最后,使用内容分析法对采集得到的中英文相关文献数据进行研判,判断这些文献是否涉及到知识融合的相关支撑理论,并运用归纳法汇总相关支撑理论,将其划分到相应的知识融合之前、知识融合过

程、知识融合之后三个阶段。

4 相关支撑理论分析

4.1 知识融合之前

大数据时代,在知识融合之前首先要考虑为什么需要进行知识融合和怎样选取融合对象。笔者通过梳理分析发现,多元表示原理(Polyrepresentation)和协同理论(Synergetics)可以从多角度解释知识融合的原因,而相关性原理(Principles of Correlation)、意义建构理论(Sense-Making Theory)和情境认知理论(Situated Cognition)可以为融合何种类型的知识解疑释惑。

(1) 多元表示原理

多元表示原理是由丹麦著名情报学家英格沃森(Peter Ingwersen)于 1992 年在《信息检索交互》(Information Retrieval Interaction)专著中提出。该原理最初是作为一种知识组织方法为提高检索效率而提出,旨在改进对知识资源的智能访问,后被视为交互式信息检索的核心认知理论成果^[31]。该原理认为,来自于不同认知的行动者在认知上有不同的表示,来自于相同认知的行动者在功能上有不同的表示。所以,多元表示原理可以解释同一信息的多种表述,有助于加深人们对信息活动的理解,也有助于指导人们的信息活动。

在大数据时代的知识表示规模大、类型多、繁杂且重复度高。不同对象的内涵特征、不同认知行动者的个体差异性、不同认知活动的情境不确定性都可能导致知识的多元表示。比如,描述不同对象的重量有不同的知识表现形式,人称为“体重”,货物称为“重量”,船称为“吨位”。当问及哪些知识与重量相关时,需要融合“体重”“重量”“吨位”为单位的全部知识以形成全面的知识体系。由于认知行动者的差异性,即使描述同一对象也有不同的知识表现形式,比如,丈夫称呼小芳为“老婆”,孩子称其为“妈妈”,父母称其为“女儿”。当问及对小芳的描述有哪些时,需要融合包含“老婆”“妈妈”“女儿”的全部表述才能形成对小芳的全面评价。描述不同情境下的同一对象也有不同的知识表现形式,比如,五一黄金周的北京故宫旅游景区客流量为“拥挤”,日常工作日的客流量为“舒适”,日常周末的客流量为“一般”。当问及北京故宫客流量多还是少时,需要融合

关于“拥挤”“舒适”“一般”的所有表述以形成客观的知识反馈。因此,多元表示原理不仅可以解释多源异构知识的繁杂表象,也可以为大数据时代的知识融合之前寻找共同的语义内容与关联知识提供支撑。

(2) 协同理论

德国斯图加特大学教授、著名物理学家哈肯(Hermann Haken)于1971年首次提出协同的概念,于1977年又系统地论述了协同理论^[32]。协同理论以系统论、信息论、控制论、突变论等为基础,主要研究不同事物的共同特征及其协同机理,着重探讨各种系统从无序变为有序时的相似性。该理论认为,系统内部各要素、子系统和要素以及子系统之间存在着各种联系,协同运作后产生的整体效用大于各部分的简单加总。

大数据时代知识融合的核心特征是要发挥多源知识的协同效果,只有多源知识融合后的整体效用大于各分散知识源的单一效用,才能更好地满足知识服务的需求。协同理论强调多种知识资源的协同整合,追求多源知识的利用及服务价值的最大化,可以解释知识融合的原因。

(3) 相关性原理

相关性原理可追溯到1901年Francis W. Parker提及的教育中各学科之间的相关性^[33]。相关性原理是系统方法论的一个基本原理,揭示了系统各要素之间以及系统与外部环境的关系。系统内部各要素之间的相互联系是有机的,各要素相互关联、相互作用、共同构成系统的整体;与外部环境有紧密联系的系统是开放系统,系统同外部环境的相关性体现在与外界发生物质、能量和信息的交换。如马费成认为任何一种情报结构都是按一定规则相互关联的,而研究和揭示情报相互关联(即相关性)的规律和规则是有效组织和检索信息、知识、情报的基础^[34]。

取代因果分析更关注相关分析是大数据时代知识分析与处理技术的深刻变革。由于相关性原理可以判定知识之间的相关性以及知识源与任务情境之间的相关性,所以,融合了相关要素、相关任务、相关情境等形成的精准知识可以为某一领域的问题求解提供更好的解决方案。

(4) 意义建构理论

布伦达·德尔文(Brenda Dervin)于1972年提出了以用户为中心的意义建构理论^[35]。该理论的核心内容是信息的不连续性、人的主体性以及情境对信息渠道和信息内容选择的影响,认为信息研究应由强调信息来源转向强调信息使用者,人不再只是被动、消极、机械的信息观察者和接受者,即认为知识是主观的、由个人建构而成,信息寻求是一种主观建构的活动,会形成不同的意义建构过程。

大数据时代的知识源结构各异、体量庞大、来源广泛、可选择性多。意义建构理论有助于揭示异构知识源对任务目标的支撑作用以及不同的知识结构对任务的理解、分析与处理。因此,人们在意义建构过程的差异性可能导致选取不同的知识源,人的主体性和知识结构的独特性又可能导致对任务有不同的理解方式与处理手段。

(5) 情境认知理论

情境认知理论出现于20世纪80年代,是由杜威(John Dewey)、维果斯基(Lev Vygotsky)、列昂节夫(Leontyev)等专家共同研究提出。该理论认为,认知过程是由情境建构、指导和支持的,个体的心理活动通常在情境中进行^[36]。认知心理学研究表明,在人类认知的整个过程中,认知都有赖于情境,我们所知觉到的意义不可能独立于对情境的解释和适应。

情境决定认知本质,情境是所有认知活动的前提。大数据时代人们对知识进行融合面临更加复杂的情境,不同的用户需求、不同的任务目标、不同的认知者等都可能以差异性的认知方式选择不同的知识源进行融合。

4.2 知识融合过程

融合过程是知识融合的关键步骤。在融合过程开始之初,如何将庞杂异构的知识转换与表示为统一规范的,使计算机可理解、组织、管理与利用的知识集合,本体论(Ontology)和形式语言理论(Formal Language Theory)可发挥重要作用。如何确定知识的细分和整合程度,量化知识颗粒度大小,为问题解决提供不同层次和不同颗粒度的知识,粒度计算理论(Granular Computing Theory)可以提供理论支撑,其中粗糙集理论(Rough Set Theory)是粒度计算理论的经典理论之一。在知识融合过程中需要根据用

户的求解需求选择合适的融合规则,由于融合知识源的不确定性以及知识性质的影响,可以基于不同的理论设计融合规则,如 D-S 证据理论(Dempster-Shafer Evidence Theory)、模糊集理论(Fuzzy Sets Theory)和知识空间理论(Knowledge Space Theory)等。

(1) 本体论

本体论源自哲学之中称为“形而上学”的分支,由 17 世纪德国经院学者郭克兰纽(Goclenius)首先使用^[37]。20 世纪 80 年代,人工智能领域的研究人员开始采用术语“本体论”来指关于模型化世界的一种理论以及知识系统的一种组件^[38]。本体就是对于共享概念体系显性的形式化表达,是一种语义数据模型和知识组织的方式,可以用来辨识和关联语义上存在对应关系的概念,并具有推理功能,有时人们也会将本体称为本体论。目前,本体广泛应用于人工智能、语义网、软件工程、图书馆学、情报学以及信息架构等领域。

在知识融合过程中,本体的运用可以解决大数据时代多知识源之间存在的语义不一致性问题,可以提供一个共享词汇表对分布式异构数据统一说明。本体能够有效支持逻辑推理,为知识融合提供语义层面的表示规则,以机器可读的方式使得知识语义能够在知识融合的过程以及在机器与人之间被传递和交流。

(2) 形式语言理论

形式语言理论是用数学方法研究自然语言(如英语)和人工语言(如程序设计语言)的产生方式、一般性质和规则的理论^[39]。形式语言是用精确的数学或机器可处理的公式定义的语言,区别于数据格式和数据结构,也不是编程语言。目前应用广泛的形式语言包括资源描述框架(RDF)、可扩展的标记语言(XML)、轻量级的数据交换格式(JSON)等。

形式语言具有高度的抽象化,可以采用形式化的手段(如专用符号、数学公式)描述语言的结构关系,并对大数据时代产生的海量非结构化知识进行统一的表示,以支持知识在融合过程中的推理。这种规范的、机器可读的标准化语言为大数据时代知识融合的实施与人机交互提供了可能。

(3) 粒度计算理论

20 世纪 90 年代后期,粒度计算理论起源

于对知识获取的简化、清晰、低成本、近似和不确定性的实际需要。粒度计算理论是一种通用的计算理论,用于有效地使用类、集群、子集、组和区间等粒度来为具有大量数据、信息和知识的复杂应用程序构建高效的计算模型^[40]。粒度是知识的基本组成单元,粒度大小是对知识抽象程度的度量,粒度化思想是以人为本从不同的粒度去思考、分析和解决问题。

在知识融合过程中,粒度计算理论可以处理大数据环境产生的复杂的、非结构化的、不完整的、不确定的知识,根据用户需求选择合适的知识颗粒度,选择既要避免整合型知识导致难以解答用户问题,也要避免细粒度知识引起问题解答效率低下的问题。

(4) 粗糙集理论

粗糙集理论由波兰数学家 Pawlak 于 1982 年提出^[41]。该理论的数据分析过程包括数据预处理、属性约简、规则提取三个阶段,旨在对不精确、不一致、不完备的知识进行分析。

粗糙集理论通过不可分辨关系将融合域中的对象进行分类,将同一类对象的集合视为一个知识,以这些已知知识为基础,可以对未知的知识进行描述,挖掘潜藏的普遍规律。在知识融合过程中,基于粗糙集理论可以对已知的知识进行分类,并对大数据环境产生的丰富的、未知的知识进行描述与归类,更好地支持问题求解。

(5) D-S 证据理论

D-S 证据理论是经典概率理论的扩展。这一理论产生于 20 世纪 60 年代,由 Dempster 于 1967 年首先提出构造不确定推理模型框架^[42],后来由其学生 Shafer 于 1976 年对这一理论进行了完善与扩充^[43]。D-S 证据理论是一种不精确推理理论,适用于决策层的知识融合问题。

D-S 证据理论有助于挖掘潜藏在大数据海洋中未知的、有深度的、有价值的知识。基于 D-S 证据理论的知识融合规则在大数据时代知识融合过程中的具体应用思路是:针对某一个命题全部可能的有意义解,即所有假设构成的集合中,对每一个推断的可能结果进行信任值计算评估,最后按照一定的判决规则选择可信度最大的假设作为融合结果。例如,基于 D-S 证据理论可以对电子商务用户数据中的

知识进行融合测算,将不确定信息转化为一个确定的区间值,最终给用户较为精准的商品推荐^[44]。

(6) 模糊集理论

1965年,美国加州大学伯克利分校电气工程系的扎德(Zadeh)教授首次提出了模糊集理论。该理论用隶属程度对中间过渡进行差异化描述,用一种精确的数学语言对模糊性进行描述^[45]。模糊集理论对解决不确定性知识融合问题具有良好的适用性,还可以与其他融合算法相结合来提高融合处理的精度,如基于模糊集理论的知识融合方法^[46]。

大数据时代,数据的爆炸导致存在大量内涵复杂、可扩展性强、价值不确定的知识,这些模糊性的知识可能相互矛盾,但对问题求解却是不可忽略的。基于模糊集理论的融合规则可以很好地处理相互矛盾的知识以及错误知识的融合,支持不同应用场景的融合需求。

(7) 知识空间理论

知识空间理论是一种有效评估知识的数学理论,起源于20世纪80年代,由心理学家Doignon和Falmagne提出^[47]。知识空间理论包括知识状态、知识结构、知识空间、学习路径等基本概念。该理论把有关特定知识领域的信息概念化为一组大型的、特定的问题,个人关于该领域的知识状态被形式化为该个人能够解决所有问题的子集,并假设任意组合的所有可能的知识状态的集合是封闭的,而满足此条件的集合族就称为知识空间。

大数据时代,数据洪流下潜藏的大量知识杂乱无章地分布在知识空间的各个角落。在基于知识空间理论的新型知识融合方法中,将知识空间中的对象结点以不同的粒度表示,且都有对应的本体和元知识集,知识对象之间的距离由知识聚合度等结构和语义关系来定量化^[30]。这种知识聚类 and 融合方法调整了空间结构,能够提高大数据时代知识搜索的策略,优化知识融合过程。

4.3 知识融合之后

知识融合之后的结果要素中可能包括新的知识源、新知识、新的知识类型、新的问题解决方法或理念、新的知识概念体系等。可视化理论(Visualization Theory)为知识融合结果的全局观提供技术支撑,有利于知识融合任务目

标的整体布局和科学决策。认知评价理论(Cognitive Evaluation Theory)和信息扩散理论(Information Diffusion Theory)可以用于对知识融合结果进行评价和反馈,有助于提高知识融合的服务质量和效率。

(1) 可视化理论

可视化理论主要是以数据作为基础,基于IT技术研究数据表示、数据处理、决策分析等一系列问题^[48]。可视化是通过数字化、虚拟化、智能化的设计将数据转换成图形或图像并对其进行交互处理的综合技术手段,能够直观、形象地将知识源和知识管理系统中的所有要素进行有效链接,实现各种资源要素的开放共享、统筹管理和高效应用。

可视化具备智能化的特点,可以直观地帮助用户在短时间内获取所需的知识资源,以丰富的呈现方式展示知识融合的成果,增加知识的动态性和易读性。而且通过直观的、可视化的图形还可以展示知识融合后的知识全局观,如本体对齐和知识图谱融合后的实体及关系概览,为大数据环境下的问题解决和科学决策提供关联分析。

(2) 认知评价理论

认知评价是指人对客观事件、事物的看法和评判,是对于情景因素支持或控制行为的一种心理评价。认知评价理论是由德西和莱恩(Deci & Ryan)于1975年提出的,又称为自我决定论。该理论认为控制行为的外部强化无视个人的自我决定,促使人们把自己的行为认知为是由外部所决定的,因此导致内在动机降低^[49]。

融合分布式异构知识源后产生的新知识可能是无效的或者无用的,甚至是不合逻辑的,所以需要对其进行评价和反馈。这一过程不仅要借助一定的算法进行自适应评价,还要通过用户的评价和反馈,因为知识融合的任务始于特定的用户需求和应用问题,用户的评价和反馈更有助于评估知识融合结果的质量。

(3) 信息扩散理论

信息扩散理论源于埃弗雷特·罗杰斯(Everett M·Rogers)提出的风险评估理论。该理论将传统的数据样本点映射成为一个模糊集合,即将单位值样本变成集值样本,从而扩展样本点的信息量进行概率评估^[50]。

基于信息扩散理论的评价机制可以提高

融合知识的准确性^[13]。例如,通常对知识融合结果的评价是由解决实际问题的用户所提供,但这样的评价反馈属于小样本,评估缺乏可靠性。对于个体来说,反馈是确定性的,但是对于所有用户来说,反馈集合是模糊性的,因此将信息扩散理论引入评估机制是可行的,有利于发现大数据时代潜在的知识,提高融合知识的语义相关性和准确性。

5 大数据时代知识融合的支撑理论架构

根据对上述相关支撑理论的系统梳理分析,基于知识融合的三阶段论,本研究设计了大数据时代知识融合的支撑理论架构。

5.1 架构设计

本研究设计的大数据时代知识融合的支

撑理论架构由三个层次、三个阶段、十五个支撑理论构成,如图 1 所示。其中,知识融合之前包括多元表示原理、协同理论、意义建构理论、相关性原理和情境认知理论等五个理论;知识融合过程包括本体论、形式语言理论、粒度计算理论、粗糙集理论、D-S 证据理论、模糊集理论和知识空间理论等七个理论;知识融合之后包括可视化理论、认知评价理论和信息扩散理论等三个理论。而且知识融合之前的多元表示原理、意义建构理论、相关性原理、情境认知理论和知识融合过程的本体论、形式语言理论还能为知识融合之后提供相关的理论支撑,知识融合过程的本体论、形式语言理论也能为知识融合之前提供相关的理论支撑。

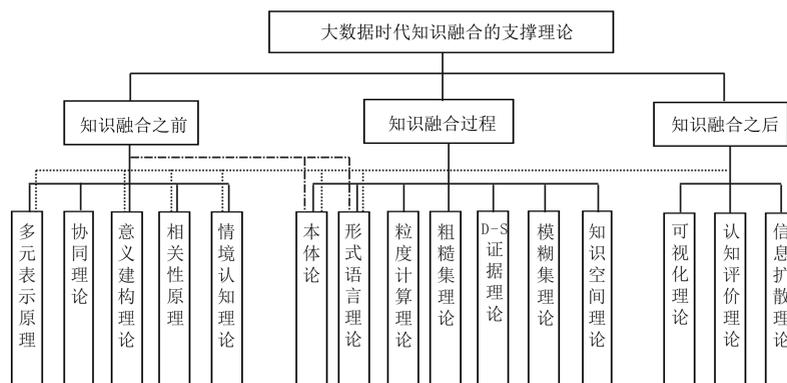


图 1 大数据时代知识融合的支撑理论架构

5.2 架构意义

本研究设计的大数据时代知识融合的支撑理论架构,从知识融合全过程视角,展示了知识融合之前、知识融合过程和知识融合之后三个阶段的相关支撑理论。该架构清晰地呈现了相关理论在知识融合三个阶段中所起到的相应的支撑作用,以及在具体的知识融合应用场景下各相关支撑理论协同作用的理论集合,表明该架构具有知识融合全过程的系统性,可以为知识融合的实现路径和服务模式提供理论指导。

在知识融合之前,多元表示原理和协同理论可以分别从知识表示的多元化和多来源异构知识源的协同整合作用解释为什么要进行知识融合。相关性原理、意义建构理论和情境认知理论可以指导选择融合对象,根据知识融合的不同情境,以人为本融合相关的知识要素

和任务,挖掘出更加全面系统的融合知识,以支持问题解决和实现知识服务价值最大化。这五个理论不仅可以对知识融合之前的多来源、跨领域、大规模的异构知识进行准确智能的发现、分析和提取提供相关理论支撑,而且多元表示原理和情境认知理论还可以运用于知识融合之后的差异化情境下的不同知识表现形式;意义建构理论还可以应用于用户对知识融合之后融合结果的评价和利用,并通过知识融合的结果不断建构和修正原有的知识结构;相关性原理还可以将知识融合之后的新知识或新的问题解决方案等运用到主题相关、要素相关、任务相关、情境相关等新问题中。

在知识融合过程中,本体论和形式语言理论可以提供机器可读的知识资源语义描述机制,有效处理语义冲突和支持知识的逻辑推理。粒度计算理论和粗糙集理论可以将知识

对象分成不同粒度层级,为问题解决过滤和筛选不同颗粒度的知识。基于 D-S 证据理论、模糊集理论和知识空间理论可以设计不同的知识融合规则和算法,满足差异化的知识需求目的。这七个理论不仅为知识融合过程中的知识转换、知识表示、知识计算和融合实施提供相关理论支撑,而且其中的本体论和形式语言理论还可以运用到对知识融合之前的知识单元和知识融合之后的融合结果进行统一的知识表示,使知识语义以一致性的形式表示、传递与人机交互。

在知识融合之后,运用可视化理论可以展示知识的全局观,有助于决策者更为直观和高效地洞悉潜在的知识、情报以及智慧。运用认知评价理论和信息扩散理论可以对知识融合之后的结果和知识融合的动态过程进行评价与反馈,以提高知识融合的效果和效率。同时,经评价和反馈之后产生的新知识还可以为新一轮的知识融合做储备,进一步解释为什么

要进行知识融合或成为后续的知识融合对象。

6 结语

知识融合是大数据时代提供精准知识服务的重要手段,是实现知识服务创新与增值的有效方式。本文设计的大数据时代知识融合支撑理论架构,不仅可以为大数据时代的知识融合全过程提供理论支撑,还可以有效指导知识融合的实现路径和服务模式。该架构由知识融合的三个层次、三个阶段、十五个理论组成,具有知识融合全过程的系统性。由于本文是通过数据的采集与归纳汇总得到的相关理论,并基于知识融合的三阶段论来设计大数据时代知识融合的支撑理论架构,所以数据集的局限性可能导致该架构难以涵盖知识融合全过程的所有支撑理论。后续研究将进一步从知识融合实现路径的角度详细分析各支撑理论处于知识融合实现路径中哪个关键步骤以及关键步骤中还缺乏哪些支撑理论。

参考文献

- [1] Preece A, Hui K, Gray A, et al. KRAFT: An agent architecture for knowledge fusion[J]. *International Journal of Cooperative Information Systems*, 2001, 10(1-2): 171-195.
- [2] 刘晓娟,李广建,化柏林.知识融合:概念辨析与界说[J]. *图书情报工作*, 2016, 60(13): 13-19, 32.
- [3] Lenat D B, Feigenbaum E A. On the thresholds of knowledge[J]. *Artificial Intelligence*, 1991, 47(1): 185-250.
- [4] Sergio R G, Alberto F, Salvador G. Fernandez A, Garcia S, et al. Big data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce [J]. *Information Fusion*, 2017, 35: 68-80.
- [5] Grebla H A, Cenan C O, Stanca L. Knowledge fusion in academic networks [J]. *Broad Research in Artificial Intelligence and Neuroscience*, 2010, 1(2): 111-118.
- [6] Smirnov A, Levashova T, Shilov N. Patterns for context-based knowledge fusion in decision support systems[J]. *Information Fusion*, 2015, 21: 114-129.
- [7] Tai C H, Chang C T, Chang Y S, Yue-Shan C. Hybrid knowledge fusion and inference on cloud environment[J]. *Future Generation Computer Systems*, 2018, 87: 568-579.
- [8] Chen C, Jafari R, Kehtarnavaz N. Improving human action recognition using fusion of depth camera and inertial sensors[J]. *IEEE Transactions on Human-Machine Systems*, 2015, 45(1): 51-61.
- [9] Andrade D, Horeis T, Sick B. Knowledge fusion using dempster-shafer theory and the imprecise dirichlet model [C]// *IEEE Conference on Soft Computing in Industrial Applications*. IEEE, 2009: 142-148.
- [10] Suzuki M, Araki D, Higashide A, et al. An approach to data fusion using uncertain knowledge in Geographical Information Systems[J]. *Electrical Engineering in Japan*, 1999, 128(4): 65-76.
- [11] Mahmoud E H, Agnes B, Serge G. Multi-source information fusion: multi-temporal images, crop growth modelling, and expert knowledge[C/OL]//2007 International Workshop on the Analysis of Multi-temporal Remote Sensing Images. IEEE, 2007. doi: 10.1109/MULTITEMP.2007.4293064.
- [12] Niu Y, Yan G, Xie G, et al. Multi-sensor fusion using knowledge-based mind evolutionary algorithm[C]//Third International Conference on Convergence & Hybrid Information Technology. IEEE Computer Society, 2008: 897-902.
- [13] Xu C, Li A, Liu X. Knowledge fusion and evaluation system with fusion-knowledge measure[C]//Second International Symposium on Computational Intelligence & Design. IEEE Computer Society, 2009: 127-131.
- [14] Preece A, Hui K, Gray A, et al. The KRAFT architecture for knowledge fusion and transformation[J]. *Knowledge -*

- based systems, 2000, 13(2-3): 113-120.
- [15] Kuo T T, Tseng S S, Lin Y T. Ontology-based knowledge fusion framework using graph partitioning[C]//Developments in Applied Artificial Intelligence, 16th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems. Springer Springer Verlag Inc, 2003: 11-20.
- [16] Dong X L, Gabrilovich E, Heitz G, et al. From data fusion to knowledge fusion[J]. Proceedings of the VLDB Endowment, 2014, 7(10): 881-892.
- [17] Bryl V, Bizer C, Isele R, et al. Interlinking and knowledge fusion[M]. Linked Open Data-Creating Knowledge Out of Interlinked Data. Springer International Publishing, 2014: 70-89.
- [18] 祝振媛, 李广建. “数据—信息—知识”整体视角下的知识融合初探——数据融合、信息融合、知识融合的关联与比较[J]. 情报理论与实践, 2017, 40(2): 16-22.
- [19] 张卫东, 左娜, 陆璐. 政府网站信息资源知识融合体系架构设计[J]. 图书情报工作, 2018, 62(17): 112-119.
- [20] 王曰芬, 岑咏华. 大数据时代知识融合体系架构设计研究[J]. 数字图书馆论坛, 2016(10): 16-24.
- [21] 徐彤阳, 邓颖慧. 基于知识融合的微信多媒体检索研究[J]. 情报科学, 2019, 37(1): 129-133, 147.
- [22] 唐晓波, 刘广超. 基于两层知识融合的金融知识服务模型研究[J]. 图书馆学研究, 2018(16): 79-85.
- [23] 罗立群, 李广建. 智慧情报服务与知识融合[J]. 情报资料工作, 2019, 40(2): 87-94.
- [24] 化柏林, 李广建. 大数据环境下多源信息融合的理论与应用探讨[J]. 图书情报工作, 2015, 59(16): 5-10.
- [25] 张心源, 邱均平. 大数据环境下的知识融合框架研究[J]. 图书馆学研究, 2016(8): 66-70, 43.
- [26] 王敬东, 李佳. 基于知识协同的大数据知识处理模型的设计[J]. 情报科学, 2015, 33(10): 140-142, 153.
- [27] 韩翔宇, 李彦, 刘勇, 等. 基于本体的航天领域知识组织方法[J]. 航天工业管理, 2018(4): 66-69.
- [28] 蒋黎黎, 梁坤, 叶爽. 基于粒度计算理论的知识融合模型研究[J]. 计算机应用研究, 2012, 29(10): 3697-3700.
- [29] 唐旭丽, 张斌, 傅维刚. 情境本体驱动的多源知识融合框架[J]. 图书情报工作, 2018, 62(22): 109-117.
- [30] 宋艳辉, 邱均平. 我国“五计学”知识融合的思考[J]. 现代情报, 2019, 39(2): 4-7.
- [31] 张新民. 多元表示与情报学[J]. 情报理论与实践, 2009, 32(7): 23-28.
- [32] Haken H. Synergetics: An introduction[M]. New York: Springer-Verlag Berlin Heidelberg, 1977.
- [33] Parker F W. Principles of correlation[J]. Course of Study, 1901, 1(6): 506-509.
- [34] 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.
- [35] Brenda D. Sense-making theory and practice: An overview of user interests in knowledge seeking and use[J]. Journal of Knowledge Management, 1998, 2(2): 36-46.
- [36] Roth W, Jornet A. Situated cognition[J]. Wiley Interdisciplinary Reviews: Cognitive Science, 2013, 4(5): 463-78.
- [37] Enskat R. Ontology[M]. Religion Past and Present. Brill Online, 2013.
- [38] 翟军. 关联政府数据原理与应用——大数据时代开放数据的技术与实践[M]. 北京: 电子工业出版社, 2016: 1-32.
- [39] 冯志伟, 周建. 自然语言形式分析的奠基力作——纪念《句法结构》出版 60 年[J]. 现代语文, 2019(1): 4-23.
- [40] Bargiela A, Pedrycz W. Toward a theory of granular computing for human-centered information processing[J]. IEEE Transactions on Fuzzy Systems. 2008, 16(2): 320-30.
- [41] Pawlak Z. Rough set[J]. International Journal of Computer and Information Sciences, 1982(11): 341-356.
- [42] Dempster A P. Upper and lower probabilities induced by multivalued mapping[J]. The Annals of Mathematical Statistics, 1967(38): 325-339.
- [43] Glenn S. A mathematical theory of evidence[M]. Princeton. NJ: Princeton University Press, 1976: 245-250.
- [44] 王越. 电子商务用户数据的知识融合研究[J]. 计算机产品与流通, 2019(5): 287.
- [45] Zadeh L A. Fuzzy sets[J]. Information & Control, 1965, 8(3): 338-353.
- [46] 周芳, 刘玉战, 韩立岩. 基于模糊集理论的知识融合方法研究[J]. 北京理工大学学报(社会科学版), 2013, 15(3): 67-73.
- [47] Doignon J P, Falmagne J C. Spaces for the assessment of knowledge[J]. International Journal of Man-Machine Studies, 1985, 23: 175-196.
- [48] 任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. 软件学报, 2014, 25(9): 1909-1936.
- [49] Deci E L. Intrinsic motivation[M]. New York: Plenum Press, 1975.
- [50] 张颖, 王美, 王静. 基于信息扩散理论的森林火灾风险分析及森林资源保护[J]. 环境保护, 2018, 46(19): 38-43.

(收稿日期: 2019-10-16)