

国外大数据管理研究热点主题可视化分析*

赵蓉英^{1,2,3}, 余波^{1,2,3}

(1. 武汉大学中国科学评价研究中心, 武汉 430072; 2. 武汉大学信息资源研究中心, 武汉 430072;

3. 武汉大学信息管理学院, 武汉 430072)

摘要: 本文以Web of Science中大数据管理研究文献作为数据源, 利用文献计量方法分析大数据管理研究现状和热点主题。运用信息可视化软件CitespaceV对所收集的文献进行学科主题词共现分析和聚类分析, 探寻隐藏在大数据管理主题研究聚类网络下有关研究热点的相关知识结构, 揭示大数据管理研究领域热点主题的发展脉络和趋势, 以期大数据管理研究在各学科领域的研究和发展提供参考。

关键词: 数据管理; 研究热点; 知识图谱; 可视化

中图分类号: G251

DOI: 10.3772/j.issn.1673-2286.2017.12.011

1 引言

“大数据”一词代表众多自主和独立来源的数据集, 具有量大、复杂和快速增长等特性。随着计算机技术、互联网技术, 特别是万维网技术的发展, 大数据在政务、医疗、商务、安全和气候等方面发挥越来越重要的作用。2008年, *Nature*出版了专刊*BigData*, 从多个方面强调海量数据带来的挑战。2011年, *Science*发布*Dealing with data*, 阐述了科学研究中大数据处理和数据科学管理等问题。2011年6月, 麦肯锡咨询公司对大数据分析、应用、管理和挑战等问题进行详细地阐述。2012年3月, 美国联邦政府宣布投资2亿多美元用以大数据研发, 并强调大数据与互联网、超级计算等具有同样重要的国际战略意义。计算硬件、网络 and 大规模并行处理框架的快速发展, 以及计算基础设施成本的降低促成大数据、云计算和物联网等计算技术已快速生成大量的数据^[1]。在大数据应用和管理方面, Zhang等介绍了大数据驱动产品生命周期管理框架所应对的挑战, 提出数据生命周期管理框架及整体数据管理的解决方案^[2]; Boone等介绍了企业已深刻认识到大数据分析的重要性, 强调改进服务部门数据管理实践的方法^[3]。

大数据的大规模效应给数据的收集、存储、管理

和应用带来极大挑战, 大数据的管理方式将面临一系列变革。大数据的管理、技术和应用存在很多争议和疑问, 需要进一步探索和创新。国内学者对大数据分析与管理进行了深入地研究, 在社会、经济各个层面都展开分析。因此, 对国外大数据分析与管理领域研究热点进行分析与探讨具有重要的意义。

2 数据获取与方法

本文数据来源于Web of Science, 检索时间为2017年7月12日。利用数据库的高级检索功能进行来源文献检索, 以“big data”和“data management”为检索词, 限定“标题”检索, 检索国外关于大数据管理研究的所有文献。通过数据清洗, 最后获得393篇有效文献作为研究的数据样本。本文对原始文献信息进行处理后, 采用文献内容分析方法结合知识图谱可视化软件CitespaceV对大数据管理研究文献进行深入分析, 构建该领域关键词知识图谱, 通过文献进行可视化处理生成研究热点网络图谱。

3 研究热点分析

关键词、主题词等代表了文献的核心与精髓, 是

* 本研究得到国家社会科学基金项目“中国学者国际学术论文影响力评价研究”(编号: 16BTQ055)资助。

对文献内容的高度概括和凝练。对关键词、主题词等进行共现分析,分析其共同出现的频次和关系,从而揭示某一学科领域的知识结构演变情况、研究热点及科学研究前沿等。利用CitespaceV构建国外大数据管理领域研究热点知识图谱。在CitespaceV初始界面选择“TimeSlicing”为所有年限,选择“Year Per Slice”以一年为一个时间段,节点类型为关键词,得到包括120个节点和282条连线的研究热点网络知识图谱(见图1)。

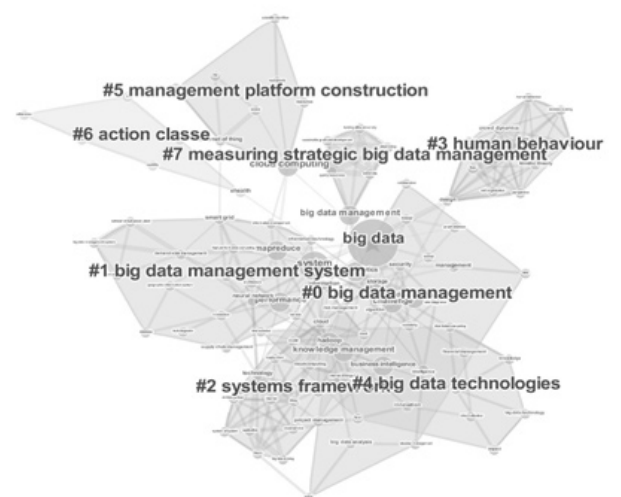


图 1 大数据管理研究热点网络知识图谱

其中圆点代表关键词,圆点大小反映了热点的强弱。圆点间的连线表示热点间的共现,连线间的密集程度代表两个主题间联系的紧密程度。通过对CitespaceV中发现集群和标签标题集群可视化操作,得到许多不规则的多边形集群。所有集群将联系紧密的关键词聚成一簇,并根据关键词的意义进行分类和标注,标注的排序代表该集群主题研究热点的程度。

通过该软件可视化后台数据networksummary-table进一步分析,得到大数据管理研究领域热点关键词的频次和突发词数据信息,剔除意义不明确的关键词后得到表1。

大数据管理领域研究热点关键词主要集中在2013—2016年。中心性最强的词为big data (0.44), system (0.26)、knowledge management (0.17) 和cloud computing (0.16) 次之。

2013年热点关键词有big data (大数据)、cloud computing (云计算)、data analytics (数据分析)、ehealth (电子健康)、home monitoring (家庭监控)、data warehousing (数据仓库)、ambient assisted living (环境辅助生活)、electronic health record (电子健康

表 1 大数据管理研究高频关键词和突发词 (部分)

频 次	中心性	关键词	年 份
150	0.44	big data	2013
20	0.26	system	2014
17	0.16	cloud computing	2013
16	0.14	big data management	2014
9	0.1	mapreduce	2013
9	0.17	knowledge management	2015
9	0.05	performance	2015
7	0.00	analytics	2016
7	0.01	data mining	2015
6	0.14	ehealth	2013
6	0.04	smart grid	2015
6	0.06	security	2014
6	0.00	ehealth	2013
5	0.06	design	2014
5	0.13	neural network	2015
5	0.17	cloud	2015
5	0.06	storage	2014
5	0.00	management	2014
5	0.01	innovation	2015
4	0.01	project management	2015
4	0.01	impact	2016
4	0.01	crowd dynamics	2016
4	0.00	emergency management	2016
4	0.00	kinetic theory	2016
3	0.00	intelligence	2016
3	0.00	information technology	2015
3	0.00	machine learning	2016
3	0.00	risk management	2016
3	0.00	big data technology	2016
3	0.00	knowledge	2016
2	0.03	electronic health record	2013
2	0.01	data warehousing	2013
2	0.00	home monitoring	2013
2	0.00	ambient assisted living	2013
1	0.00	supply chain management	2016
1	0.00	business intelligence	2015

记录)等。这反映了国外在2013年的研究主要集中在大数据技术及其应用管理的发展阶段,大数据凸显了数据分析的重要性和紧迫性。其中,国外大数据在电子健康记录方面应用广泛,充分体现数据分析的便捷和高效。

2014年热点关键词有system(系统)、big data management(大数据管理)、storage(存储)、security(安全性)、management(管理)、design(设计)等。通过这些热点关键词发现,国外在该领域的研究开始遇到数据安全和管理方面的挑战,进一步重视大数据的安全性和管理。

2015年热点关键词有knowledge management(知识管理)、business intelligence(商业智能)、data mining(数据挖掘)、smart grid(智能电网)、neural network(神经网络)、cloud(云)、project management(项目管理)、innovation(创新)、information technology(信息技术)等。这些高频关键词体现了大数据管理与信息技术的融合,表明信息技术对数据管理的重要性。

2016年热点关键词有analytics(分析)、kinetic theory(动力学理论)、crowd dynamics(人群动态)、knowledge(知识)、risk management(风险管理)、emergency management(应急管理)、impact(影响)、intelligence(情报)、supply chain management(供应链管理)、machine learning(机器学习)、big data technology(大数据技术)、cloud computing(云计算)等。随着大数据形式的多样化和信息技术的发展,大数据管理已渗透到社会经济、医疗、安全等各个行业。

通过2013—2016年的热点关键词可以发现,大数据分析等一直是该领域研究的热点,这充分体现出大数据管理与数据分析密不可分。同时还发现数据管理与云计算、数据挖掘及知识管理密切相关。

CitespaceV可通过中介中心性来衡量和发现文献的重要性,中心度高的节点通常代表网络中重要的节点。通过关键词聚类 and 主题分析有利于发现该领域研究的重点和热点。结合关键词的中心性深入探测热点主题,最终得到big data management(#0)、big data management system(#1)、system framework(#2)、human behavior(#3)、big data technologies(#4)、management platform construction(#5)、action classes(#6)、measuring strategic big data management(#7) 8个主题。

为进一步分析热点主题,本文在对国外该领域知识图谱分析的基础上,结合文献调研结果,对该领域8个热点主题进行整合后得到的热点主题为大数据分析与管理系统(整合#0和#1)、大数据与信息技术(整合#2和#4)、大数据安全与隐私管理(整合#3和#5)、大数据行业应用与管理(整合#6和#7)。

3.1 大数据分析与管理系统

目前,国外大数据分析与管理系统领域研究主题较多。大数据的信息源呈现出自主性,并出现多样化异构数据,如社交媒体数据、网络开放数据、传感器数据等。多源异构数据对采集、存储、处理、分析和管理带来前所未有的挑战。分布式处理技术是数据分析的有效途径。国外政府和相关机构通过无线网络与云计算分析和处理大数据,研究云计算环境中大数据聚合与高效管理的相关性来提高数据的分析和处理效率^[4]。

大数据分析与管理系统结合本体论、语义处理、云计算等视角从大量数据集中挖掘有价值的数据来进行科学分析与处理。另外,数据建模是大数据分析与管理系统的创新。在大数据环境下,数据管理需要通过原始方法解决数据复杂性建模和架构问题,利用建模和开发系统技术来处理大数据的管理模式^[5]。对于大数据的数据捕获、数据存储、数据分析和数据可视化等复杂性问题需要建立新的管理和服务模式来提升利用效率,同时,大数据建模、管理系统和可视化等将提升数据管理策略^[6]。目前,研究人员通过调查大数据在管理、金融、通信等领域的影响,以探寻更好的大数据分析与管理系统的方法^[7]。

3.2 大数据与信息技术

大数据管理离不开信息技术,通过改进技术以及大数据数量、速度和多样性管理的特点,以提供大数据集成技术的上下文感知智能服务。通过信息技术来提升数据管理,Gil等提出基于本体、语义链接和云计算的多种架构综合智能服务^[8]。

通过强调存储、预处理、处理和安全性来调查大数据管理的可行技术,数据技术仍然需进一步提升^[9]。MapReduce等分布式处理技术在移动设备中的应用可分析出系统故障,阻止使用移动设备应用程序的处理。Park等提出一种基于利用率和移动速率的分组技术和阶段分组方法,可有效减少组管理的开销^[10]。因此,通过设计分类法来分析这些技术的关键方面,总结大数据管理技术的不同层面来提高数据技术和管理。

3.3 大数据安全与隐私管理

在大数据时代,数据安全和隐私管理一直是国内外关注的焦点。数据安全涉及经济、社会、文化等各个

领域。

国外基于安全云计算的数据框架体现在智能电网的大数据信息管理,在现代能源基础设施方面发挥关键作用。智能电网通过安全数据技术创新,提升电力服务效率。基于安全云计算的智能电网大数据信息管理框架^[11]除了为信息管理和大数据分析提供不同类型的计算服务外,还提供身份的加密、签名和代理重新加密等安全服务。

3.4 大数据行业应用与管理

大数据已进入行业管理和应用层面,大数据管理是企业系统/平台的发展趋势。社会各行业都可以通过大数据的应用和管理来提高企业效益,尽可能降低企业成本。研究和分析大数据在运营或供应链管理中的应用,辅助战略运作策略的制定^[12]。

4 结语

首先,通过对国外大数据管理领域热点主题的可视化研究,得到该领域研究的热点关键词,包括大数据、云计算、大数据管理与分析、数据仓库、电子健康记录、知识管理、商业智能、数据挖掘、数据安全和隐私、大数据技术、供应链管理等。结合文献调研,可发现大数据管理领域研究正关注大数据分析 with 知识管理的结合^[13]。

其次,通过热点关键词聚类分析发现,该领域研究的热点主题大致可分为大数据分析与管理、大数据与信息技术、大数据安全与隐私管理、大数据行业应用与管理。这4类主题涉及政务、医疗、商务、安全和气候等多个方面。大数据分析和应用面临新的机遇与挑战,其处理方法根据不同层面的信息需求在不断深化和更新。同时,大数据管理也将随着信息技术和互联网技术发展而不断改进数据管理方式,以此满足不同层面团体和个人的需求。

最后,大数据管理研究将继续呈现多元化研究方向。时间性是大数据管理的新研究方向,已经开发的时间关注模型不适合处理大数据的新特征,大数据管理方法和建议亟待进一步更新^[14]。从数据质量管理和数据使用经验理论角度考虑,企业对大数据分析的要求是提升企业数据质量^[15]。

国外大数据管理领域研究正处于多元化发展和不

断深化的过程中,该领域引起不同行业和学界的高度重视,特别是政务、商务、医疗、安全和气候等行业的广泛关注和研究;随着互联网技术的更新和大数据的不断累积,大数据管理有待进一步探索和完善;大数据分析和管理将在不同领域开拓新的应用空间,为各领域提供更好的数据管理和服务平台。

参考文献

- [1] GIL D, SONG I Y, ALDAND J F, et al. Big Data. New approaches of modelling and management[J]. Computer Standards & Interfaces, 2017, 54(11): 61-63.
- [2] ZHANG Y, REN S, LIU Y, et al. A framework for big data driven product lifecycle management[J]. Journal of Cleaner Production, 2017, 159(8): 229-240.
- [3] BOONE C A, SKIPPER J B, HAZEN B T. A framework for investigating the role of big data in service parts management[J]. Journal of Cleaner Production, 2017, 153(1): 687-691.
- [4] ORIKE S, BROWN D. Big data management: an investigation into wireless and cloud computing[J]. International Journal of Interdisciplinary Telecommunications and Networking, 2016, 8(4): 34-50.
- [5] GIL D, SONG I Y. Modeling and management of big data: challenges and opportunities[J]. Future Generation Computer Systems, 2016, 63(10): 96-99.
- [6] GIL D, TRUJILL O, SONG I Y. Big data trends: modelling, management and visualization[J]. Expert Systems, 2016, 33(4): 362-363.
- [7] OHSHIMA N, KESHAVARZ H, WAN H H, et al. Study of open issues on big data management[C]// Proceedings of the 11th International Conference on Innovation and Management. [S.l.]: [s.n.], 2014: 1245-1246.
- [8] GIL D, FERRÁNDEZ A, MORAMORA H, et al. Internet of things: a review of surveys based on context aware intelligent services[J]. Sensors, 2016, 16(7): 1069.
- [9] SIDDIQA A, HASHEM I A T, YAQOOB I, et al. A survey of big data management: taxonomy and state-of-the-art[J]. Journal of Network & Computer Applications, 2016, 71(8): 151-166.
- [10] PARK J S, KIM H, JEONG Y, et al. Two-phase grouping-based resource management for big data processing in mobile cloud computing[J]. International Journal of Communication Systems, 2014, 27(6): 839-851.
- [11] ALELAIWI A. A collaborative resource management for big IoT data processing in Cloud[J]. Cluster Computing-The Journal of Networks Software Tools and Applications, 2017, 20(2): 1791-1799.
- [12] ADDO-TENKORANG R, HELO P T. Big data applications in operations/supply-chain management: a literature review[J]. Computers & Industrial

- Engineering,2016,101(11):528-543.
- [13] HOTA C,UPADHYAYA S,AI-KARAKI J N.Advances in secure knowledge management in the big data era[J].Information Systems Frontier,2015,17(5):983-986.
- [14] CUZZOCREA A.Temporal aspects of big data management: state-of-the-art analysis and future research directions[J].International Symposium on Temporal Representation and Reasoning,2016,22(1):180-185.
- [15] KWON O,LEE N,SHIN B.Data quality management, data usage experience and acquisition intention of big data analytics[J].International Journal of Information Management,2014,34(3):387-394.

作者简介

赵蓉英,女,1961年生,博士生导师,研究方向:信息计量与科学评价,E-mail:zhaorongying@126.com。
余波,男,1981年生,博士研究生,研究方向:信息计量与科学评价。

Visualization Analysis of Hot Topics of Big Data Management in Foreign

ZHAO RongYing^{1,2,3}, YU Bo^{1,2,3}

(1.Research Center for Chinese Science Evaluation, Wuhan University, Wuhan 430072, China; 2.Center for Studies of Information Resources, Wuhan University, Wuhan 430072, China; 3.School of Information Management, Wuhan University, Wuhan 430072, China)

Abstract: This paper takes big data management research literature in the Web of Science database as the data source, using bibliometric methods to analyze the current situation and hot topics of big data management. Using the information visualization software CitespaceV to carry on the subject co-occurrence and the cluster analysis, explores the relevant knowledge structure of research hotspots that are hidden in big data management subject research cluster networks, revealing the trend of hot topics in big data management research field, so as to provide a reference for the research and development of big data management research in various disciplines.

Keywords: Data Management; Research Hotspots; Knowledge Maps; Visualization

(收稿日期: 2017-10-31)

■ 书 讯 ■

《科技报告体系构建研究》

为推进我国科技报告制度建设,强化科技报告资源共享服务,贺德方研究员率领中国科学技术信息研究所科技报告研究团队,进行了国家社会科学基金重点项目“中国科技报告资源体系构建”(11ATQ006)研究,并对20多年来中国科学技术信息研究所相关研究和实践进行了归纳、凝练、整理和补充,最终形成了《科技报告体系构建研究》。

本书作为国家社会科学基金重点项目的研究成果,总结了科技报告产生发展的管理历程、凝练了科技报告制度的建设路径、制订了科技报告资源的整合方案,提出了科技报告体系的构建模式,归纳了科技报告实践的操作过程。本书对各级科技计划管理人员强化科技计划项目过程管理具有借鉴作用,对科研人员撰写高质量科技报告具有指导作用,对各类科研机构做好科技报告呈交、推进科技项目的规范管理和机构知识库建设具有参考价值,对图书信息机构做好科技报告深层次加工和收藏利用具有引导作用,也可供高校信息管理、科技政策与管理等专业研究生学习参考。

《科技报告体系构建研究》于2014年12月由科学技术文献出版社出版,定价78.00元。