

# 数据科学在国内管理学研究中的应用综述

李志国<sup>1,2</sup> 钟 将<sup>3</sup>

(重庆工商大学长江上游经济研究中心 重庆 400067)<sup>1</sup> (重庆工商大学管理学院 重庆 400067)<sup>2</sup>  
(重庆大学计算机学院 重庆 400044)<sup>3</sup>

**摘 要** 基于大数据时代的新学科——数据科学的研究方法正在被包括管理学在内的其他学科应用。首先,探讨了以大数据为理论基础的数据科学研究范式与管理学研究的经典范式之间的联系和区别。其次,分析了国家自然科学基金委认定的 A 类重要管理学期刊文献及引证文献,对当前国内管理学领域为数不多的基于数据驱动的公共管理、基于复杂网络仿真的网络行为管理和基于多源数据融合的创新管理等热点领域进行了分类梳理。然后,归纳总结了当前国内管理学领域采用数据科学研究方法的特征。最后,提出了数据科学在管理学科学研究应用中的趋势,即范式融合、大数据利用、场景融合、专家合作。

**关键词** 数据科学,管理学,大数据,数据驱动

中图分类号 G304 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.09.005

## Application of Data Science in Management Science Study: State-of-the-art in Domestic

LI Zhi-guo<sup>1,2</sup> ZHONG Jiang<sup>3</sup>

(Research Center for Economy of Upper Reaches of Yangtse River, Chongqing Technology and Business University, Chongqing 400067, China)<sup>1</sup>  
(School of Management, Chongqing Technology and Business University, Chongqing 400067, China)<sup>2</sup>  
(College of Computer Science, Chongqing University, Chongqing 400044, China)<sup>3</sup>

**Abstract** As a new discipline, data science which is based on the era of big data is applied to other disciplines including management. Firstly, the connections and differences between the scientific research paradigm based on data science and the classical paradigm of management research were proposed. Secondly, the relative literatures and citation literatures in the major Management Journal of A-class identified by the National Natural Science Foundation of China were identified, and the current management research hotspots like data-driven based public management, network behavior management based on complex network simulation and innovation management based on multisource data fusion were classified and summarized. And then, the characteristics of data science used in current management field were summarized. Finally, the trend of the application of data science in management science research was proposed, that is paradigm fusion, big data utilization, scene fusion and expert cooperation.

**Keywords** Data science, Management science, Big data, Data-driven

1998年,图灵奖获得者 Jim Gray 把以大数据及其相关技术为支撑的数据密集型科学研究方法称为继实验科学、理论科学、计算科学后的“第四范式”(the fourth paradigm)<sup>[1]</sup>。尽管科学史和科学哲学史上尚未有过“第几范式”这样的分类,哲学界也对此有诸多讨论<sup>[2]</sup>,但数据科学作为大数据时代的一门学科已经成为不争的事实<sup>[3-4]</sup>。作为数据科学基础的大数据,近年来被广泛应用于电子商务、政府公共服务、医疗服务等领域,并取得了巨大的经济和社会效益,这不仅激励了计算机科学领域内大量科学工作者对大数据理论、方法和技术的持续深入研究,也引发了社会科学工作者的高度关注,促进

了数据科学相关融合交叉学科的发展<sup>[5]</sup>。2012年, *Harvard Business Review* 中刊出的“Big Data: The Management Revolution”一文指出,正是因为有了大数据,管理者才可以通过测量更加深刻地认识自己的业务,并将这些认知用于改善决策和商业表现<sup>[6]</sup>。这一观点迅速在商业界和管理学界引起了轰动,极大地推动了大数据方法在管理学研究中的应用。

大数据来源于物理世界和人类社会<sup>[7]</sup>。构成人类社会的基本单元是“人”和“组织”,它们正是现代管理学的研究对象,加之与实证主义范式在统计分析上存在共同特征,在管理学研究中采用大数据研究方法,因而具有重要的理论意义、巨大

收到日期:2018-03-12 返修日期:2018-05-24 本文受长江经济带绿色低碳可持续发展研究项目(950617005),国家重点研发计划(2017YFB1402400)资助。

李志国(1977—),男,博士,高级工程师,主要研究方向为数据科学、战略管理、精准招商、区块链等, E-mail: lizhiguo@ctbu.edu.cn(通信作者);  
钟 将(1974—),男,教授,博士生导师,主要研究方向为数据挖掘、可信计算机系统、服务计算。

的实践价值和鲜明的时代特性。

本文第1节先从科学方法论的视角探讨以大数据为基础的数据科学研究方法与管理学经典研究方法的区别和联系;第2节从国内管理学重要权威期刊文献的分析入手,总结、提炼和分析管理学学科研究应用数据科学的热点领域;第3节提出国内管理学科领域采用数据科学研究中方法的特点;第4节研判数据科学范式在管理学科学研究应用中的趋势;最后总结全文。

## 1 数据科学范式与管理学经典范式的联系与区别

现代科学中的大多数学科都采用以波普(Karl R. Popper)为代表的后实证主义作为判断研究方法是否科学的依据,其主要的标准是“可证伪性”,认为仅仅依靠归纳总结现象规律得到的理论不是真正的科学<sup>[8]</sup>。Thomas Kuhn以海王星的科学发现为例(天王星的轨道不满足哥白尼的日心说并不是因为日心说理论有错误,而是因为存在一颗我们尚未发现的行星,即海王星),指出即使观察与理论不一致,也并不一定代表理论错误,进而提出科学研究的“范式转移”理论<sup>[9-10]</sup>。“范式”可以理解为某个科学领域在一定时期被公认的思维方式和开展科学研究的一般过程或方法。不少学科的发展历程表明,使用新的范式取代旧的范式是一个极其艰难的历史过程。

### 1.1 管理学研究的经典范式

从方法论视角看,管理学的主流科学研究范式有两种<sup>[11-12]</sup>:1)以统计分析、比较分析、案例分析为代表的实证主义范式,采用“现象描述—理论假设—证明理论”的归纳法论证思想;2)以数学公式推导为代表的解释主义范式,采用“现象描述—数学抽象—理论推导”的演绎法论证思想。实证主义范式和解释主义范式都是从管理学现象入手,通过构建理论模型来解释现象,然后对提出的理论进行证明,以期将该理论外推,从而解释更多的同类型现象或对其发展作出预测。

从研究目的视角看,管理学的经典研究方法可以划分为3种<sup>[13]</sup>:1)描述型研究方法,主要回答“是什么”的问题,即“实然”问题,如市场调查、经济普查、民意测验等;2)解释型研究方法,主要回答“为什么”的问题,探索现象背后的支配逻辑,如因果研究、相关分析、理论构建等;3)规范型研究方法,主要回答“该怎样”的问题,即“应然”问题,如资政研究、咨询建议、对策研究等。

管理学的经典研究范式与自然科学和其他人文社会科学有着较大的区别。自然科学的研究以可重现的实验和严密的数学或逻辑推导为主,人文社会科学则以比较、归纳演绎、因果分析、结构与功能分析为代表的思辨方法为主<sup>[14]</sup>,而管理学的研究方法既有逻辑思维也有思辨思维。我们观察到的管理学现象要么影响因素太多,要么环境受限,因此不同的学者描述和刻画这些现象的数据存在较大差异,从而导致不同学派往往在观点上较难达成一致,以库恩为代表的不少学者认为管理学的研究范式更像宗教而不是科学。

### 1.2 数据科学范式与管理学研究经典范式的联系

数据科学范式与管理学研究范式的重要联系主要来源于以下几个方面。

1)数据科学的学科内涵。数据科学本身是一门交叉学科,以领域知识、统计学和机器学习等相关理论为基础<sup>[15-16]</sup>,其知识体系(见图1)包括理念、方法、技术和工具等理论,以及数据加工(Data Wrangling)、数据管理、数据计算、数据分析和数据可视化等专业技术知识<sup>[17]</sup>。其中,领域知识的多样性决定了数据科学应用场景的多样性,管理学也是重要的领域之一,因此数据科学与管理学应用场景存在交叉联系。

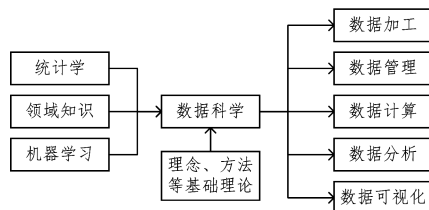


图1 数据科学的知识体系

Fig. 1 Knowledge system of data science

2)二者在统计分析方法上的一致性。管理学实证主义范式经常利用统计分析方法,例如对理论构建提出的相关概念进行的问卷调查、抽样数据的回归分析、因子分析和结构方程模型构建等。与数据科学的统计分析方法一样,管理学实证主义的统计分析方法也关注数据或者变量之间的相关性<sup>[18]</sup>。模型的基本形式是 $y=f(x,\beta,\epsilon)$ ,其中 $y$ 是因变量, $x$ 是自变量, $\beta$ 是参数, $\epsilon$ 是误差(扰动项), $f$ 是相关函数。

3)对数据价值的关注。数据科学和计算机科学关于大数据的研究的基本问题是计算复杂性问题,核心思想是如何对Volume, Variety, Velocity和Value(即大数据的“4V”特征)的数据进行高效快捷的处理,从而发掘数据价值。数据价值也是管理学所关注的焦点,无论是抽样统计数据、问卷调查数据、单案例或多案例分析数据,其根本目的都是挖掘出这些数据背后的理论意义和现实指导意义<sup>[19]</sup>。

### 1.3 数据科学范式与管理学研究经典范式的区别

数据科学范式与管理学研究范式的根本区别主要体现在以下几个方面。

1)研究的起点和逻辑顺序。管理学研究范式从典型的现象观察和分析中利用先验知识进行理论模型构建或者理论假设,然后用有限的抽样数据进行统计分析和假设检验,从而证明所提理论的解释力,并用于认识和解决现实问题。而数据科学范式是在没有模型和理论假设的情况下直接对大样本数据甚至全样本数据进行分析,然后用数据分析结果来认识和解决现实问题,例如基于语料库统计技术路线的机器翻译<sup>[20]</sup>。这也正是《Wired》杂志主编Anderson发出“理论已终结”这一惊人论断的原因<sup>[21]</sup>,尽管这一观点并未得到普遍认同。

2)研究的分析方法。管理学研究范式强调变量之间的因果关系,即强调找到 $y=f(x,\beta,\epsilon)$ 中的 $\beta$ ,侧重于解释型分析,关注理论价值。数据科学范式更强调变量之间的相关关系,对海量数据进行搜索、比较、聚类和分类,模型可以简化为

$y=f(x,\epsilon)$ ,直接呈现变量之间的相关关系,侧重于预测型分析,更加关注预测的商业价值<sup>[6]</sup>。

## 2 数据科学在国内管理学研究的热点应用领域

管理学学界普遍认同现代管理学起源于“科学管理之父”泰勒于1911年出版的《科学管理原理》一书。管理学作为一门学科,近年来在国内取得了长足发展,形成了纷繁复杂的学科丛林。本文理想的研究方法是从管理学学科体系出发,采用描述性方法对各领域的数据科学研究方法进行总结分析(这也是管理学常用的规范分析研究范式),但后续研究发现目前运用数据科学研究方法的国内管理学相关文献还不足。因此,本文的研究方法利用数据科学中常用的“相关分析”方法,从文献检索到的特定权威文献全样本数据出发,利用领域专家知识对管理学热点研究问题进行分类判断,再辅以文献引证网络进行拓展,重点对数据科学和管理学研究相关文献进行总结和分析。

### 2.1 研究热点分析

本文从国家自然科学基金委认定的22种A类重要管理学期刊出发,以“数据科学”和“大数据”为关键词对这些期刊文献题目和关键词进行全面搜索,共检索到至2017年末的相关文献75篇,去除征稿、机构介绍等明显不相关的文献后剩余71篇,文献数量不多。

根据文献的研究领域和出版时间,本文采用德尔菲法征集多名专家意见,对71篇文献进行了分类,如表1所列。需要特别指出的是,计算科学、经济学、情报学和统计学并非管理学的研究范畴,但是这些文献研究与管理学研究场景高度相关,因此本文也对其进行了统计呈现。由表1可知,为数不多的文献在2013年—2017年期间呈现逐年上升的态势,文献数量较多的研究领域主要有公共管理、复杂网络、创新管理、财务管理与企业社会责任、营销管理、数据治理等管理学热点研究领域;结合数据科学的视角,以基于数据驱动的公共管理、基于复杂网络仿真的网络行为管理、基于多源数据融合的创新管理为主。以这些文献为基础,结合其引证文献,本文分以下几个领域进行归纳分析。

表1 文献分类统计描述

Table 1 Statistical description of literature classification

文献研究领域	(单位:篇)					
	2013	2014	2015	2016	2017	小计
公共管理(灾害、健康、扶贫等)			4	2	7	13
网络行为管理	1	4	2	3	1	11
创新管理	2	1	1	2	4	10
财务管理与社会责任	0	2	0	1	1	4
营销管理	1		1		1	3
数据治理			2		1	3
决策管理		1		1		2
情报学相关				4	5	9
经济学相关	1	1		1		3
计算科学相关		1	1			2
统计学相关				1		1
其他			2	5	3	12
小计	5	10	13	20	23	71

### 2.2 基于数据驱动的公共管理热点领域

大数据能够促进政府的整体数据分析和运用水平,从而

有效提升公共服务能力,已经成为各级政府和学界的共识。国发[2015]50号文《促进大数据发展行动纲要》指出,建立“用数据说话、用数据决策、用数据管理、用数据创新”的管理机制,利用大数据“能够揭示传统技术方式难以展现的关联关系”这一特性,通过高效采集、有效整合、深化应用等资源整合促进数据融合,从而为政府有效处理复杂社会问题提供新的手段。管理学界沿着数据驱动管理效能提升这一指导思想,用数据科学的理念和方法针对精准扶贫、灾害管理、健康管理等公共管理热点问题开展了不少探索性的研究。

在公共管理领域的精准扶贫问题研究上,潘竟虎和胡艳兴受到徐康宁等利用DMSP-OLS(美国国防气象卫星计划,由美国国家海洋和大气管理局对外提供相关产品)全球夜间灯光数据来测算中国实际经济增长率<sup>[22]</sup>的启发,使用DMSP-OLS夜光影像来剔除NPP-VIIRS(美国国家极轨环境卫星)影像中的异常值,构建了由灯光指数和多维贫困指数构成的贫困精准识别模型,并开展了基于栅格和县域尺度的贫困空间识别<sup>[23]</sup>。在具体方法上通过线性回归寻找平均夜间灯光指数和多维贫困指数之间的关联关系。章昌平和林涛认为大数据对精准扶贫具有重要的支撑作用,并构建了以贫困人口为中心的大数据管理整合方案理论框架模型<sup>[24]</sup>,方法上采用了规范分析的研究范式。汪磊等<sup>[25]</sup>同样采用规范分析的方法,认为利用大数据能够减少甚至消除扶贫机制中的信息不对称现象,并分析了大数据4V特征与扶贫逻辑之间的耦合机制,进而构建了精准扶贫大数据管理信息平台理论框架。

在灾害管理方面,罗江华梳理了我国灾害数据库建设的基本情况<sup>[26]</sup>,包括中国科学院的人地系统主题数据库、农业部的自然灾害数据库、国家地震局的地震科技文献数据库和地震灾情描述数据库,以及相关的科学数据共享工程;分析了西部灾害大数据建设面临的问题,并提出了灾害大数据统一规划、资源共享、重视数据文化元素等大数据建设的发展思路。庞素琳<sup>[27]</sup>针对巨灾风险大数据信息流庞杂、数据量巨大、种类繁多、关联性复杂的特点,利用规范分析方法和算法设计思想,提出了巨灾风险大数据应急分类和分解的算法,以期解决巨灾大数据有效挖掘、归类、分拣和处理的问题。

在健康管理方面,吴友富等认为大数据会颠覆公共健康服务领域,但同时也面临隐私风险,并在此背景下利用规范研究方法为政府部门在数据治理、技术标准和法律法规制定方面的政策提出了建议<sup>[28]</sup>。徐曼等从异构实体数据融合、多模态数据管理、数据分治3个方面对大数据支持下的精准医疗决策成果进行了归纳分析,并指明了互联网环境下智能医疗的发展方向<sup>[29]</sup>。郭旦怀等利用来自哨点医院、食品和互联网3个方面的数据,建立食源性疾事件探测模型,并进一步引入人口、交通、食品生产等大数据进行风险预测,研究中采用了大数据聚类方法来选择症状和食物信息对病人进行聚类<sup>[30]</sup>,利用动态上下文挖掘微博数据,从而确定事件窗口。

### 2.3 基于复杂网络仿真的网络行为管理热点领域

李国杰院士极富洞见地指出,关系网络是大数据背后的共性问题<sup>[7]</sup>,数据网络是复杂关联大数据的独特存在形式,因

而复杂网络分析是数据科学这个“第四范式”的重要基石。管理学的研究对象即“人”与“人”的关系、“人”与“组织”的关系、“组织”与“组织”的关系,可以利用关系网络来刻画,而关系网络在数字空间中的映射就是数据网络的一种存在形式,因而不少管理学者利用复杂网络上的数据传播、搜索、聚类研究思路,开展了社交网络运营管理和网络舆情管理等网络行为管理场景的研究。

社交网络运营管理研究方面,主要采用实证主义研究范式将大数据用于关联性分析,以期提升社交网络的运营效率。例如,冉晓斌等<sup>[31]</sup>以激发社交网络用户活跃度为导向,建立了对活跃用户行为理解的个体活跃度概念模型,以及个体网络规模、同伴活跃度、关系强度、性别、年龄 5 个参数与个体活跃度关联关系的 5 个研究假设。与问卷调查等传统抽样统计方法通常为数百条实证数据不同,他们采用了某在线社交网络平台中 120 余万用户的大样本数据。研究中利用 Tobit 模型对被解释变量与解释变量之间的关系进行回归分析,并印证了所建立的个体活跃度概念模型的正确性。

网络舆情管理方面,主要采用解释主义范式的研究方法把大数据作为理论模型的仿真研究输入,以期为舆情控制提供理论支撑。例如,张峰等提出了一个基于依从、趋同和内化 3 种状态决策偏移概念的舆论演化动力学模型<sup>[32]</sup>。该模型定义社会网络  $G=(V,E)$  和观点集合  $S(t)$ , 构建了节点  $v_i$  在时刻  $t$  与相邻节点交互收到的群体一致性压力函数  $C(i,t)=v_i * \sum_{k \in N^+} A(k)$  及偏移牵引力函数  $D(i,j,t)=\mu * \cos(s_i(t) - s_j(t)\pi/2)$ , 建立依从、趋同和内化的节点状态及状态转移策略,并在 BA 无标度网络模型上进行了仿真研究。孙毅等从百度提供的 140 个最常用的与通货膨胀相关的关键词中选出 14 个,利用主成分分析法将多个网络搜索关键词合成为有限个指标,绘制了网络搜索行为的通胀预期指数<sup>[33]</sup>。徐映梅和高一铭也利用百度搜索热度指标,提出了一种基于门限回归的低频舆情指数构建方法<sup>[34]</sup>来辅助预测 CPI, 实证其可以领先官方数据 40~65 天。米子川和李毅将改进的捕获移出模型应用到社交网络传播抽样估计中,并利用新浪微博数据对北京“雾霾”事件的舆情传播进行了实证分析<sup>[35]</sup>。

#### 2.4 基于多源数据融合的创新管理热点领域

创新是引领发展的第一动力,党的十九大报告提出加快建设创新型国家,《国家创新驱动发展战略纲要》多处强调大数据在科技创新、制度创新、管理创新、商业模式创新等领域发挥着重要作用。数据科学的预测性分析方法和强大的数据处理技术能够分析多源数据且具有更强的时效性,因此受到创新决策相关研究的重视。

技术创新管理方面,朱东华等整合了“目标驱动决策”和“数据驱动决策”理念,提出了大数据环境下面向技术创新的双向决策模型<sup>[36]</sup>。周潇等提出了一个基于多源数据和配套 SAO 理论框架的识别技术创新路径模型<sup>[37]</sup>,该模型利用文本挖掘的方法,通过主题词簇合并进行主题提取,并对生物医药代表性技术的固体脂质纳米粒子(SLN)进行了实证分析。该实证分析从 WOS(Web of Science)及 Medline 数据库、DII

(Derwent Innovations Index) 专利数据库以及 ABI/Inform (Abstracts of Business Information) 商业数据集超过 10 万条数据中先分离出与 SLN 相关的 3000 多条数据,然后利用 Goldfile 软件工具获取 SAO 结构,再用德尔菲法进一步聚焦,最后邀请领域专家共同确定技术创新路径。多源数据技术创新路径的基本思路如图 2 所示。

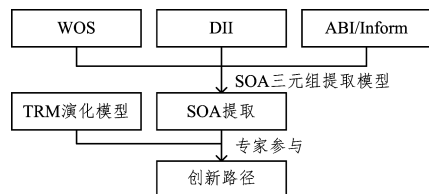


图 2 多源数据技术创新路径

Fig. 2 Innovation path of multi-source data

胡小君和张誉宁以近十年来 DII 专利数据库收录的中美高校授权专利为大数据分析样本,构建技术多样性指数来测度技术多样性,分析两国高校在 8 个行业中的专利分布及其多样性与差异性<sup>[38]</sup>。郭峰等建立了大数据环节下企业行为对创新能力和绩效的作用机理理论模型,采用问卷研究的实证主义范式对 252 家具有大数据特征的企业进行实证检验<sup>[39]</sup>。李天柱等利用案例研究方法研究了以数据和软件为主的新一代信息产业中存在的接力创新现象<sup>[40]</sup>,认为“第四范式”是催生接力创新的根本力量。

营销模式创新方面,金晓彤等证明了联动式数据库营销模式具有数据整合功能、价值共创导向、利益相容和共享 3 方面特征,从而构成联动式数据库营销的基本框架<sup>[41]</sup>,并利用一汽大众的单案例研究方法进行了实证。商业模式创新方面,田敬等设计了一种适用于跨区域多业态零售行业的大数据及商业智能系统框架体系<sup>[42]</sup>,并介绍了 HDBI 系统的关键技术、实现方法以及在国大 36524 便利店的商业实践案例。

管理模式创新及决策创新方面,徐辉提出了绩效数据信息化、平台开放化、分析综合化、评估可触化、过程可预测化、管理精细化、反馈实时化等基于大数据的人员绩效管理特征,并在此基础上提出了公共部门管理的创新模式<sup>[43]</sup>。

#### 2.5 基于机器学习等方法的其他管理学领域

为数不多的“第四范式”管理学研究除了相对集中在公共管理、网络行为管理、创新管理等这 3 个领域外,在其他领域也有少量的文献,使用的数据科学方法包括支持向量机、决策树等机器学习方法,涉及的研究方法有规范研究和实证研究。

例如,客户关系管理方面,牟刚和袁先智通过收集 SAP 运营数据和客户财务数据,使用 Logistic 回归方法研究了企业对客户的信用评级,把大数据作为模型的实证数据<sup>[44]</sup>;秦学志和李静一分别利用国内某银行的 51770 笔和 34535 笔交易流水的大数据样本集<sup>[45]</sup>,采用支持向量机(SVM)模型建立异常账户的识别机制,分辨率达到 85% 以上。大数据治理方面,郑大庆等从治理目标、权力层次、对象及范围、实际问题 4 个维度来阐述大数据治理的概念,并提出了大数据治理的逻辑框架<sup>[46]</sup>;蒋玉石等总结了在大数据背景下网络行为定向广告(Online Behavioral Advertising, OBA)与网络消费者用户

隐私需要关注的问题和相关研究<sup>[47]</sup>。项目管理方面,杨青等构建了基于大数据的工程项目管理时间维度、领域维度、方法维度的三维结构模型,探讨了大数据在项目管理方面的应用前景<sup>[48]</sup>。上市公司股东行为方面,王红武等利用国泰安数据库和CCER数据库的60万条记录数据,把决策树算法用于股东统计分类和行为分析上,分析了股东持股排名与减持行为之间的关联性<sup>[49]</sup>。企业社会责任方面,沈弋等认为大数据推动企业社会责任信息披露从一维到多维的嬗变、从言方行圆到言行一致的根基演化、从报告到报道的路径演化<sup>[50]</sup>。军事管理方面,胡晓峰等指出作战体系分析中网络化体系能力评估在深度挖掘方面需要关注大数据方法,在理论分析上需要关注复杂网络方法<sup>[51]</sup>。

### 3 数据科学在国内管理学研究中的应用特点

对数据科学和管理学研究范式的联系和区别进行分析,并对国内管理学科权威期刊上发表的文献进行梳理。从整体上看,数据科学在国内管理学研究领域的应用呈现出以下几个显著的特点。

1)数据科学在管理学领域的学术应用明显滞后于产业实践。迄今为止,利用数据科学和大数据方法开展管理学科领域研究的理论文献还屈指可数,大部分文献还处于利用数据科学的理念、通过管理学规范研究方法阐释大数据“应该”用于领域研究,属于回答“实然”问题。相比而言,产业界更加注重实用主义,不过多纠缠于数据的理论解释,偏好于走“数据直接到价值”的捷径。例如,近年来国内以BATJ为代表的互联网企业利用大数据、人工智能等技术相继推出了语音识别、人脸识别、客服机器人、翻译机等系列产品,对人们的生产和生活已经产生了深刻的影响<sup>[52-53]</sup>。在取得巨大的经济效益的同时,这些产业实践也重构了战略管理、组织管理、流程管理、客户关系管理、绩效管理等多项管理学研究的热点场景。然而,在这些场景的管理学问题研究中,如何利用数据科学研究范式还处于初步探索阶段。

2)数据科学的个别领域专业数据和方法已经开始形成新的“范式”并渗透到管理学相关领域,如复杂网络上的数据传播分析、夜间灯光数据和搜索引擎数据的利用。这些领域的相关研究在国外开展得较早,也相对比较成熟,并且都取得了轰动的社会影响效果,因而受到国内学者的重视和借鉴。比较有代表性的有:Google Flu Trends利用复杂网络分析在2009年提前几周成功预测了H1N1的爆发<sup>[54]</sup>;Henderson利用夜间灯光数据测量经济增长<sup>[55]</sup>,现实复杂网络存在的小世界网络<sup>[56]</sup>特性和无标度特性<sup>[57]</sup>等。

3)数据科学中的大样本数据分析方法与管理学研究范式相融合,并取得了一定的效果。大数据带来的“大样本”数据正在拓展管理学以问卷调查为代表的“小样本”数据用于实证研究和解释研究,从而推动了以计算机仿真为理论解释方法在管理学理论证明中的应用。传统问卷调查的样本数量通常为数百份,而大数据所提供的“大样本”数据动辄以数万、数百万计(例如前文所述的在线社交网络平台的120余万用户的

大样本数据<sup>[31]</sup>、ABI/Inform商业数据集超过10万条的数据<sup>[37]</sup>),丰富和拓展了管理学理论观念测量、变量关联关系回归等实证研究。

### 4 数据科学在国内管理学研究中的应用趋势

从计算机科学的视角看,近年来围绕计算复杂性这个基本问题的研究在多源异构大数据感知、采集、存储、分析、可视化等大数据底层支撑计算领域取得了长足的进展,为下一步管理学等交叉学科利用“第四范式”开展学术研究奠定了基础<sup>[58]</sup>。但是,数据科学的基础大数据分析技术也面临稀疏性、大噪声的挑战,因而带来了管理学实证研究范式中变量(或构念)测量、变量之间关联关系、因果判断等经典统计学理论方法在假设检验、置信度方面的困惑。同时,在没有模型和假设的情况下就对大数据开展研究,与管理学解释性研究范式、规范性研究方法之间存在逻辑冲突,通过大数据聚类分析得到的结论究竟是反映了管理学研究对象的一种现象,再以现象出发发展理论,还是直接反映了研究对象的规律和理论,用于指导我们对现象的认识?在这两大挑战的基础上,数据科学在国内管理学研究中的应用趋势可以总结为以下几点。

1)数据科学范式将进一步融入管理科学研究范式中。尽管人类的科学技术发展历程表明研究范式的转换是极其困难的,但是大数据时代背景下大规模数据的泛在性、实时性和高速传播性,使得管理学界能够以前所未有的方式从多个维度对研究对象进行刻画,不得不面对传统理论难以合理解释现实世界的困惑,从而推动数据科学范式进一步融入管理学研究范式之中。

2)管理学研究中如何利用大数据是热点问题。与传统管理学研究场景中的数据相比,大数据具有多源异构、实时性、交互性等新的特征。例如,对于知识管理<sup>[59]</sup>中难以用二维逻辑展示的非结构化数据、UGC(User Generated Content)等<sup>[60]</sup>,如何处理和利用这些数据不仅需要数据科学领域在算法和技术方面有进一步的突破,同时也要求管理学领域研究方法有一定转变,从而成为前沿研究问题。

3)数据科学与管理学场景全面融合。作为数据科学的基础,大数据正在重构生产和生活方式,管理学决策、组织、领导、控制、创新的职能正在被重新定义,大数据与管理学科的融合将呈现多学科、多场景的特征,如在公共管理、商务管理、决策管理等细分学科领域,以及农业、制造业、医疗健康、教育、智慧城市等行业领域管理学问题及场景研究中<sup>[61-62]</sup>,这一趋势已逐步显现。

4)数据专家和领域专家合作解决预测性分析和解释性分析矛盾。针对数据科学预测性分析和解释性分析的矛盾,预测性分析可以由不具备管理领域知识的数据专家完成,然后由管理领域专家对预测结果进行解释性分析,从而形成“先数据科学范式,后管理学解释性范式”的协作研究模式<sup>[63]</sup>。

**结束语** 从2008年*Nature*发表Big Data专刊<sup>[64]</sup>、2011年*Science*发表Dealing with data专刊<sup>[65]</sup>以来,基于大数据和

相关技术的“第四范式”研究方法正逐步引起计算机学界和管理学学界的重视。2016 年开始,教育部陆续批准了部分高等院校开设“数据科学与大数据技术”专业;党的十九大报告提出推动互联网、大数据、人工智能和实体经济的深度融合,将数据科学、大数据及相关技术推到了时代的前沿。在这样的时代背景下,本文分析了作为物理社会、人类社会和数字空间共性科学问题的数据科学与管理学经典研究范式的区别和联系,梳理了国内管理学应用数据科学研究的热点领域,讨论了当前数据科学在国内管理学研究的应用特点,研判了数据科学在管理学研究中的趋势。当前国内管理学界采用大数据研究方法的文献成果还寥寥无几,数据科学和管理学的交叉融合还有待进一步探索和研究。

**致谢** 感谢重庆大学经济与工商管理学院赵骅教授给予本文的指导,以及重庆大学马克思主义学院黄向前副教授、重庆交通大学经济与管理学院龚科教授给予本文的启发;同时感谢参与文献热点分类的各位专家学者的专业建议。

### 参 考 文 献

- [1] HEY T, TANSLEY S, TOLLE K. The Fourth Paradigm: Data-intensive Scientific Discovery[M]. Microsoft Research, 2009.
- [2] DONG C Y, XUE Y H. Critical Discussion on Data-intensive, Big Data & The Fourth Paradigm[J]. Studies In Dialectics of Nature, 2017, 33(5): 74-80, 86. (in Chinese)  
董春雨, 薛永红. 数据密集型、大数据与“第四范式”[J]. 自然辩证法研究, 2017, 33(5): 74-80, 86.
- [3] CHAO L M, XING C Y, ZHANG Y. Data Science Studies: State-of-the-art and Trends[J]. Computer Science, 2018, 45(1): 1-13. (in Chinese)  
朝乐门, 邢春晓, 张勇. 数据科学研究的现状与趋势[J]. 计算机科学, 2018, 45(1): 1-13.
- [4] DAVENPORT T H, PATIL D J. Data scientist; the sexiest job of the 21st century[J]. Harvard Business Review, 2012, 90(10): 70-76.
- [5] AL-JARRAH O Y, YOO P D, MUHAIDAT S, et al. Efficient Machine Learning for Big Data [J]. Big Data Research, 2015, 2(3): 87-93.
- [6] MCAFEE A, BRYNJOLFSSON E. Big Data: The Management Revolution[J]. Harvard Business Review, 2012, 90(10): 60-68, 128.
- [7] LI G J, CHENG X Q. Research Status and Scientific Thinking of Big Data[J]. Bulletin of the Chinese Academy of Sciences, 2012, 27(6): 647-657. (in Chinese)  
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, 27(6): 647-657.
- [8] POPPER K R. Conjectures and Refutations: The Growth of Scientific Knowledge[J]. Economica, 1962, 15(61).
- [9] KUHN T S. The Structure of Scientific Revolutions(Second Edition)[J]. Clio, 1970(7): 554-555.
- [10] KUHN T S. In The Structure of Scientific Revolutions[J]. American Journal of Physics, 1970, II(7): 14-16.
- [11] RUE L W, BYARS L L. 管理学技能与应用(第 11 版)[M]. 北京: 北京大学出版社, 2005.
- [12] 周三多. 管理学——原理与方法(第 5 版)[M]. 上海: 复旦大学出版社, 2009.
- [13] WU J X, YU Y D. Discussion on the methodology of case study in management research[J]. Studies in Science of Science, 2004, 22(z1): 105-111. (in Chinese)  
吴金希, 于永达. 浅议管理学中的案例研究方法——特点、方法设计与有效性讨论[J]. 科学学研究, 2004, 22(z1): 105-111.
- [14] 吴鹏森, 房列曙. 人文社会科学基础(第 2 版)[M]. 上海: 上海人民出版社, 2008.
- [15] SCHUTT R, O'NEIL C. Doing Data Science; Straight Talk from the Frontline[M]. O'Reilly Media, Inc., 2013.
- [16] MORAES R M D, LÓPEZ L M. Computational Intelligence Applications for Data Science[J]. Knowledge-Based Systems, 2015, 87(6): 1-2.
- [17] MATTMANN C A. A vision for data science[J]. Nature, 2013, 493(7433): 473-475.
- [18] WEI J R, JIANG P. The Statistical Connotation of Data Science [J]. Statistical Research, 2014, 31(5): 3-9. (in Chinese)  
魏瑾瑞, 蒋萍. 数据科学的统计学内涵[J]. 统计研究, 2014, 31(5): 3-9.
- [19] CHEN H, CHIANG R H L, STOREY V C. Business intelligence and analytics; from big data to big impact[J]. Mis Quarterly, 2012, 36(4): 1165-1188.
- [20] XING Y K, MA S P. A Survey on Statistical Language Models [J]. Computer Science, 2003, 30(9): 22-26. (in Chinese)  
邢永康, 马少平. 统计语言模型综述[J]. 计算机科学, 2003, 30(9): 22-26.
- [21] ANDERSON C. The End of Theory; The Data Deluge Makes Scientific Method Obsolete[J]. Wired Magazine, 2008, 16: 7-8.
- [22] XU K N, CHEN F L, LIU X Y. The Truth of China Economic Growth: Evidence from Global Night-time Light Data[J]. Economic Research Journal, 2015(9): 17-29. (in Chinese)  
徐康宁, 陈丰龙, 刘修岩. 中国经济增长的真实性: 基于全球夜间灯光数据的检验[J]. 经济研究, 2015(9): 17-29.
- [23] PAN J H, HU Y X. Spatial Identification of Multidimensional Poverty in China Based on Nighttime Light Remote Sensing Data[J]. Economic Geography, 2016, 36(11): 124-131. (in Chinese)  
潘竟虎, 胡艳兴. 基于夜间灯光数据的中国多维贫困空间识别[J]. 经济地理, 2016, 36(11): 124-131.
- [24] ZHANG C P, LIN T. Habitat Simulation; Integration of Poverty Population-centric Big Data in Targeted Poverty Alleviation[J]. Journal of Public Management, 2017(3): 124-134. (in Chinese)  
章昌平, 林涛. “生境”仿真: 以贫困人口为中心的大数据关联整合与精准扶贫[J]. 公共管理学报, 2017(3): 124-134.
- [25] WANG L, XU L, WANG X. The Coupling Analysis and its Mechanism Innovation of China's Accurate Poverty-relief Operation Mechanism in the Context of Big Data—Taking Guizhou and Gansu as Examples[J]. Journal of Public Management, 2017(3): 135-143. (in Chinese)

- 汪磊,许鹿,汪霞.大数据驱动下精准扶贫运行机制的耦合性分析及其机制创新——基于贵州、甘肃的案例[J].公共管理学报,2017(3):135-143.
- [26] LUO J H. Big data construction and problem analysis of disaster management in Western China[J]. Management World, 2015(4):176-177. (in Chinese)
- 罗江华.西部地区灾害管理的大数据建设与问题探析[J].管理世界,2015(4):176-177.
- [27] PANG S L. The algorithms of emergency classification, decomposition and sorting for processing a catastrophe risk of large data and its applications[J]. Systems Engineering—Theory & Practice,2015,35(3):743-750. (in Chinese)
- 庞素琳.巨灾风险大数据处理应急分类、分解、分拣算法与应用[J].系统工程理论与实践,2015,35(3):743-750.
- [28] 吴友富,万岩,范静,等.大数据时代健康信息隐私管理的政府行为研究[J].管理世界,2017(1):174-175.
- [29] XU M, SHEN J, YU H Y. A Review on Data-driven Healthcare Decision-making Support[J]. Industrial Engineering and Management,2017,22(1):1-13. (in Chinese)
- 徐曼,沈江,余海燕.数据驱动的医疗与健康决策支持研究综述[J].工业工程与管理,2017,22(1):1-13.
- [30] GUO D H, CUI W J, GUO Y C, et al. Foodborne disease event detection and risk assessment based on big-data[J]. Systems Engineering—Theory & Practice, 2015, 35(10):2523-2530. (in Chinese)
- 郭旦怀,崔文娟,郭云昌,等.基于大数据的食源性疾病事件探测与风险评估[J].系统工程理论与实践,2015,35(10):2523-2530.
- [31] RAN X B, LIU Y W, JIANG J H. Big Data Analysis of the Active Behavior in Social Network: The Perspective of Network Externality[J]. Management Sciences in China, 2017, 30(5):77-86. (in Chinese)
- 冉晓斌,刘跃文,姜锦虎.社交网络活跃行为的大数据分析:网络外部性的视角[J].管理科学,2017,30(5):77-86.
- [32] ZHANG F, WU B, WANG B, et al. Decision offset based on dynamics opinion model[J]. Systems Engineering—Theory & Practice,2014,34(S1):172-178. (in Chinese)
- 张峰,吴斌,王柏,等.基于决策偏移的舆论演化动力学模型[J].系统工程理论与实践,2014,34(S1):172-178.
- [33] SUNY, LV B F, CHEN H, et al. Measurement and application of inflation expectations from the perspective of big data[J]. Management World,2014(4):171-172. (in Chinese)
- 孙毅,吕本富,陈航,等.大数据视角的通胀预期测度与应用研究[J].管理世界,2014(4):171-172.
- [34] XU Y M, GAO Y M. Construction of the Public Opinion Index of CPI based on the Internet Big Data[J]. The Journal of Quantitative & Technical Economics,2017(1):94-112. (in Chinese)
- 徐映梅,高一铭.基于互联网大数据的CPI舆情指数构建与应用——以百度指数为例[J].数量经济技术经济研究,2017(1):94-112.
- [35] MI Z C, LI Y. Capture-Removal Model Sampling Estimation Based on SNS-oriented Big Data[J]. Application of Statistics and Management,2016,35(3):424-434. (in Chinese)
- 米子川,李毅.面向SNS大数据的捕获移出模型抽样估计[J].数理统计与管理,2016,35(3):424-434.
- [36] ZHU H D, ZHANG R, WANG X F, et al. Research on the Methodology of Technology Innovation Management with Big Data[J]. Science of Science and Management of S. & T., 2013, 34(4):172-180. (in Chinese)
- 朱东华,张巍,汪雪峰,等.大数据环境下技术创新管理方法研究[J].科学学与科学技术管理,2013,34(4):172-180.
- [37] ZHOU X, HUANG L, MA T T. A research on the technology innovation pathway identification model from the perspective of big data[J]. Science Research Management, 2017, V38(10):1-9. (in Chinese)
- 周潇,黄璐,马婷婷.大数据视角下的技术创新路径识别研究[J].科研管理,2017, V38(10):1-9.
- [38] HU X J, ZHANG Y N. A comparative study on the characteristics of technological diversity in patents between Chinese universities and American universities[J]. Science Research Management,2017,38(11):75-81. (in Chinese)
- 胡小君,张誉宁.中美高校创新群体技术多样性特征比较研究[J].科研管理,2017,38(11):75-81.
- [39] GUO F, ZOU B, GUO J Y, et al. Research on the Mechanism of Firm Behavior Effect on Innovation Capability and Firm Performance in the Context of Big Data[J]. Science of Science and Management of S. & T., 2017,38(4):126-136. (in Chinese)
- 郭峰,邹波,郭津毓,等.大数据环境下企业行为对创新能力与企业绩效的作用机理研究[J].科学学与科学技术管理,2017,38(4):126-136.
- [40] LI T Z, MA J, FENG W. The relay innovation in the information technology industry: a multiple case study[J]. Studies in Science of Science,2014,32(3):447-453. (in Chinese)
- 李天柱,马佳,冯薇.信息产业的接力创新:一个多案例研究[J].科学学研究,2014,32(3):447-453.
- [41] JIN X T, WANG T X, YANG X. Research on the Construction of Linkage-type Database Marketing Model in Big Data Area A Case into FAW[J]. China Industrial Economy, 2013(6):122-134. (in Chinese)
- 金晓彤,王天新,杨潇.大数据时代的联动式数据库营销模式构建——基于“一汽大众”的案例研究[J].中国工业经济,2013(6):122-134.
- [42] TIAN Q, WANG S Y, E E J, et al. Design, implementation and practice of big data and business intelligence system for retail industry[J]. Systems Engineering—Theory & Practice, 2017, 37(5):1282-1293. (in Chinese)
- 田歆,汪寿阳,鄂尔江,等.零售大数据与商业智能系统的设计、实现与应用[J].系统工程理论与实践,2017,37(5):1282-1293.
- [43] XU H. Performance Improvement and Management Model Innovation of Public Sector Personnel Based on Big Data[J]. China Soft Science,2017(1):50-58. (in Chinese)
- 徐辉.基于大数据的公共部门人员绩效提升与管理模式创新[J].中国软科学,2017(1):50-58.
- [44] MOU G, YUAN X Z. Empirical study for enterprise internal

- credit rating under big data framework[J]. *Journal of Systems Engineering*, 2016, 31(6): 808-815, 849. (in Chinese)
- 牟刚,袁先智. 大数据架构下企业内部信用评级的实证研究[J]. *系统工程学报*, 2016, 31(6): 808-815, 849.
- [45] QING X Z, LI J Y. Abnormal Account Monitoring Method for Bank Loans using Big Data Samples[J]. *Systems Engineering Theory Methodology Applications*, 2016, 25(4): 677-682. (in Chinese)
- 秦学志,李静一. 基于大数据样本的银行异常账户监测方法[J]. *系统管理学报*, 2016, 25(4): 677-682.
- [46] ZHENG D Q, HUANG L H, ZHANG C H, et al. Concept and Reference Architecture of Big Data Governance[J]. *R&D Management*, 2017, 29(4): 65-72. (in Chinese)
- 郑大庆,黄丽华,张成洪,等. 大数据治理的概念及其参考架构[J]. *研究与发展管理*, 2017, 29(4): 65-72.
- [47] JIANG Y S, ZHANG H Y, JIA J, et al. Research on behavior oriented advertising(OBA) and consumer privacy concerns under the background of big data[J]. *Management World*, 2015(8): 182-183. (in Chinese)
- 蒋玉石,张红宇,贾佳,等. 大数据背景下行为定向广告(OBA)与消费者隐私关注问题的研究[J]. *管理世界*, 2015(8): 182-183.
- [48] YANG Q, WU G N, WANG L Z. Big data: A new perspective of the engineering project management driven by data[J]. *Systems Engineering—Theory & Practice*, 2017, 37(3): 710-719. (in Chinese)
- 杨青,武高宁,王丽珍. 大数据:数据驱动下的工程项目管理新视角[J]. *系统工程理论与实践*, 2017, 37(3): 710-719.
- [49] WANG H W, ZHU S T, CAI H B. Study on the Shareholders' Behaviors in Listed Companies Based on Decision Tree Methods[J]. *Application of Statistics and Management*, 2017, 36(1): 139-150. (in Chinese)
- 王红武,朱绍涛,蔡海博. 基于决策树算法的上市公司股东行为研究[J]. *数理统计与管理*, 2017, 36(1): 139-150.
- [50] SHEN Y, XU G H, WANG Z Y. Information disclosure of corporate social responsibility consistent with words and deeds—evolution framework in big data environment[J]. *Accounting Research*, 2014(9): 29-36. (in Chinese)
- 沈弋,徐光华,王正艳. “言行一致”的企业社会责任信息披露——大数据环境下的演化框架[J]. *会计研究*, 2014(9): 29-36.
- [51] HU X F, ZHANG Y, LI R J, et al. Capability evaluating problem of networking SoS[J]. *Systems Engineering—Theory & Practice*, 2015, 35(5): 1317-1323. (in Chinese)
- 胡晓峰,张显,李仁见,等. 网络化体系能力评估问题[J]. *系统工程理论与实践*, 2015, 35(5): 1317-1323.
- [52] MENG X F, CI X. Big Data Management: Concepts, Techniques and Challenges[J]. *Journal of Computer Research and Development*, 2013, 50(1): 146-169. (in Chinese)
- 孟小峰,慈祥. 大数据管理:概念、技术与挑战[J]. *计算机研究与发展*, 2013, 50(1): 146-169.
- [53] LI R, SU C, JIA J, et al. Analysis of Development Status of World Artificial Intelligence Based on Scientific Measurement[J]. *Computer Science*, 2017, 44(12): 183-187. (in Chinese)
- 李悦,苏成,贾佳,等. 基于科学计量的世界人工智能领域发展状况分析[J]. *计算机科学*, 2017, 44(12): 183-187.
- [54] COOK S, CONRAD C, FOWLKES A L, et al. Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A(H1N1) Pandemic[J]. *Plos One*, 2011, 6(8): e23610.
- [55] HENDERSON J V, STOREYGARD A, WEIL D N. Measuring Economic Growth from Outer Space [J]. *American Economic Review*, 2012, 102(2): 994-1028.
- [56] STROGATZ S H, STROGATZ S H. Exploring Complex Networks [J]. *Nature*, 2001, 410(6825): 268-276.
- [57] JEONG H, TOMBOR B, ALBERT R, et al. The large-scale organization of metabolic networks[J]. *Nature*, 2000, 407(6804): 651-654.
- [58] ZENG R X, WANG L, DU H T. Research of Knowledge Network and Structure of Big Data Research from the Perspective of Knowledge Management[J]. *Journal of The China Society for Scientific and Technical Information*, 2016, 35(11): 1173-1184. (in Chinese)
- 曾润喜,王琳,杜洪涛. 基于知识管理视角的大数据研究网络与结构研究[J]. *情报学报*, 2016, 35(11): 1173-1184.
- [59] LI Z G, FENG Y, ZHONG J, et al. Distributed Knowledge Management Model Based on Super-P2P[J]. *Computer Science*, 2007, 34(7): 184-186. (in Chinese)
- 李志国,冯永,钟将,等. 基于 Super—P2P 的分布式知识管理模型[J]. *计算机科学*, 2007, 34(7): 184-186.
- [60] NOV O. What motivates Wikipedians? [J]. *Communications of the Acm*, 2007, 50(11): 60-64.
- [61] FENG Z Y, GUO X H, ZENG D J, et al. On the research frontiers of business management in the context of Big Data[J]. *Journal of Management Sciences in China*, 2013, 16(1): 1-9. (in Chinese)
- 冯芷艳,郭迅华,曾大军,等. 大数据背景下商务管理研究若干前沿课题[J]. *管理科学学报*, 2013, 16(1): 1-9.
- [62] XU Z B, FENG Z Y, GUO X H, et al. Big-data driven management and decision-making Frontiers[J]. *Management World*, 2014(11): 158-163. (in Chinese)
- 徐宗本,冯芷艳,郭迅华,等. 大数据驱动的管理与决策前沿课题[J]. *管理世界*, 2014(11): 158-163.
- [63] FAN J, FANG H, HAN L. Challenges of Big Data analysis[J]. *National Science Review*, 2014, 1(2): 293-314.
- [64] FRANKEL F, REID R. Big Data: Distilling meaning from data [J]. *Nature*, 2008, 455(7209): 30.
- [65] STAFFS. Dealing with data Challenges and opportunities Introduction[J]. *Science*, 2011, 331(6018): 692.
- [66] JIANG Q, FU Y, LI W S, et al. Research on Predicate—Based Sampling Technology of Big Data[J]. *Journal of Chongqing Institute of Technology*, 2017, 31(8): 120-124. (in Chinese)
- 姜群,傅瑜,李文生,等. 基于谓词的大数据抽样技术研究[J]. *重庆理工大学学报*, 2017, 31(8): 120-124.