

【理论·探索】

厚数据研究综述*

●李成熙 文庭孝 (中南大学 长沙 410013)

【摘要】文章通过回顾厚数据研究的理论起源和发展,利用 WOS 和 CNKI 检索相关文献,梳理国内外厚数据研究现状,剖析和解读研究中的前沿热点和问题,为后续研究提供支持和建议。当前国内外厚数据相关研究主要集中在厚数据定义研究、厚数据应用场景研究和厚数据分析方法研究等方面,未来应继续围绕基础研究问题进行深入探索和改进,并尝试在厚数据融合研究、厚数据方法研究和厚数据应用研究等方面不断进行拓展。

【关键词】厚数据 大数据 厚描 定性分析 数据故事化 可视化

【中图法分类号】G250.7

【文献标识码】A

【文章编号】1003-7845(2022)01-0008-07

【引用本文格式】李成熙,文庭孝.厚数据研究综述[J].高校图书馆工作,2022(1):8-14

1 引言

人类活动持续不断地产出海量数据,其中所蕴含的数据价值成为社会发展的重要情报来源,数据则成为工业时代迈向数字时代的重要资源和生产要素,大数据借助于计算机的强大算力和分析力,被广泛应用在数据密集型产业领域,大数据分析成为当下最炙手可热的定量分析方法之一。无论是科学研究、企业竞争还是政府决策都越来越信任和依赖大数据。人类受益于科技发展带来的变化,也逐渐习惯于利用数据解决问题,倾向于“用数据说话”的定量分析方法去揭示问题的相关关系,但大数据背后揭示事物因果关系的定性分析似乎却被暂时遗忘。

在大数据与人工智能时代,人类更应该正面思考对大数据定量分析方法的过度依赖,回归数据意义本身,更好地挖掘数据背后的故事和因果关系,正如人工智能先驱、贝叶斯网络之父、图灵奖获得者 Pearl 教授所说“若求大智慧,当知真因果”。因此,如何有效开发利用厚数据,引起了业界和学术界的关注。现有文献表明,国内外厚数据研究主要集中在厚数据定义研究、厚数据应用场景研究和厚数据分析方法研究等方面。

2 大数据与厚数据

2.1 大数据

大数据(Big Data, BD)概念在20世纪80年代萌芽,随后不断被关注和提及,其内涵不断丰富、外

延不断扩大,尽管各行业对大数据一词的定义并未达成统一,但现有定义中对大数据的描述都体现出一个共同特征,即大数据是指超过了我们现有数据处理能力的海量数据集。大数据也呈现出许多与小数据时代的信息不同的独有特征,得到各行业普遍认可的有大数据“4V”特征,即 volume(容量大)、variety(种类多)、velocity(速度快)和 value(价值密度低)^[1]。大数据的“4V”特性以及大数据时代相较于小数据时代数据运用的差异而言,大数据分析更倾向于要总体而不要抽样、要效率而不要精准、要相关而不要因果。随着信息技术的发展和人类对数据的深入挖掘,科学研究不可避免地需要借助数据技术,利用定量分析研究方法实现其方法与成果的科学化与精准化。大数据分析为人类社会带来全新的发展契机无可否认,但若持续加深对大数据的依赖,难免落入大数据傲慢与偏见的“陷阱”中,导致社会发展被视一切为数据并以数据为中心的“数据主义”世界观支配和控制^[2]。因此,厚数据相关研究开始受到业界关注。

2.2 厚数据

在大数据强势的时代背景下,对大数据背后的隐性意义和价值的挖掘成为新的研究趋势。厚数据研究的产生和兴起也有其特定的社会背景,时代发展和数据资源的广泛应用及其现实价值为厚数据研究提供了动力。厚数据概念来源于人类学的“厚描(深描)”,1990年人类学家克利福德·格尔茨提出

* 本文系国家社会科学基金重大招标项目“基于大数据的科教评价信息云平台构建和智能服务研究”(19ZDA348)研究成果之一。

厚描述的概念:“厚描述不仅包括事实,还包括评论和解释,并具有特定细节、概念结构和丰富含义。”^[3]但厚数据一词作为在大数据时代背景下与大数据对比分析的使用,最早出现在Wang T(2013)《大数据需要厚数据》一文中,她结合人类学“厚描(深描)”概念提出了“厚数据”,即“用民族志方法来解释在大数据视觉化与分析背后的意义”^[4]。

自Wang T提出厚数据的概念后,国内外许多学者开始关注到这个在大数据时代背景下衍生出来的新词,并对其进行了较为深入的剖析解读。Steven Maxwell指出人们过度沉迷于数据信息的量,却忽略了“质”的部分,也就是分析法所能揭示的商业洞察。但数据量越大并不意味着生成的洞察和创新就一定越多。知名咨询顾问公司ReD负责人Madsbjerg(2014)在《华尔街日报》发表题为《“厚”数据的力量》的文章,指出厚数据可以协助企业理解消费者在接触产品与服务时产生的情感以及内在的脉络,因此更能协助企业面对瞬息万变的商业挑战^[5]。除了商业领域从业者的理解外,国内学者刘琳等(2015)则提出,厚数据是指那些丰富的、具有前后关联性的数据,它们只能通过丢开电脑、深入实际生活才能获得^[6]。叶丽雅(2015)认为厚数据是通过小样本分析,找出特定人群的需求。大数据呈现的是数据,厚数据呈现的是故事。大数据揭示的是“是什么”,厚数据展现的是“为什么”^[7]。赵兴龙(2017)提出所谓的厚数据,是和大数据相对而言的,指在某个特定教育情景中用于反映学生全面发展的质性数据,包括学生的兴趣、爱好等,媒介载体包括文本、视频、故事、图片等,重在揭示学生成长过程中的情感、态度、某种经历甚至对某种活动的意义等^[8]。贾男等(2017)则认为厚数据是指样本数量不大但涵盖的信息维度众多的数据^[9]。孙智中等(2020)指出,Mannik和McGarry借用厚描述的概念给出了厚数据的过程定义,厚数据的产生借助于对行为的观察,并在参与性观察的基础上探索特定的社会背景和文化^[3]。克里斯蒂安·马兹比尔格(2020)在其著作《意会:算法时代的人文力量》中借用人类学家克利福德·格尔茨提出的“深描”(thickdescription)概念,将大数据时代难以言传、只能意会的信息称为厚数据^[10]。

从上述众多厚数据定义中不难看出,国内外学者在剖析解读厚数据定义时,都以大数据作为理论基础和分析对象,强调了厚数据蕴含的隐性价值,尤其指出要充分挖掘和利用厚数据背后的情感和意

义。可见,大数据与厚数据之间紧密的逻辑关联和价值互联。此外,在大数据研究过程中,学者们还提出了许多衍生概念,并与厚数据作对比。基于厚数据与薄数据的区别,Ang(2019)将厚数据定义为通过对用户进行观察或互动获得的丰富信息,包括用户的价值观、目标和消费行为^[11]。武涛等(2015)从大数据概念出发构建了“数据谱系图”,对慢数据、快数据、厚数据、轻数据等进行了深入解析,尤其强调了厚数据区别于大数据的关键在于是否充分考虑“人”的因素,实质在于厚数据将数字理性和人类感性有机结合,并指出厚数据等于“大数据+情景分析”^[12]。袁曦临(2017)则认为薄数据呈现的是事实,而厚数据呈现的包括事实及其前后联系和意义,薄数据应该向厚数据转化,方法则是依靠情景化分析^[13]。尽管学界尚未对厚数据的概念形成权威一致的共识,但通过对国内外研究现状的梳理可以发现,目前对“厚数据的理论基础来源于人类学哲学思想”和“厚数据是定性分析方法”这两个观点已基本达成共识。

3 厚数据的应用场景

在国内外现有文献和实践应用中,厚数据应用场景多元,较为集中在情报学、教育学、公共管理、医疗健康、市场营销和社区改造等多个领域。

3.1 厚数据在情报学领域的应用

情报学作为研究信息、数据、情报和知识的重要学科,对数据的研究是其应有之义,厚数据在情报学领域的应用体现在对于厚数据理论与方法的研究。袁曦临(2017)从中美智库研究与发展的差异出发,提出智库的研究性取决于多元化的跨界研究与专题化的研究资源建设,而厚数据则成为其揭示事实背后的因果联系和核心意义的重要方法^[13]。Bor-nakke等(2018)强调大数据与厚数据存在互补关系,以此提出“大一厚数据混合”,通过两个案例证明了如何从异构的“大数据—厚数据”分析中获得洞察力,并列举了四种常见的混合策略及其用法^[14]。Fiaidhi等(2019)利用厚数据分析范式,着重阐述了厚数据分析在社交媒体中识别对话社区的重要性^[15]。孙智中等(2020)基于大数据分析局限,将厚数据引入情报学信息分析过程,提出厚数据分析的研究模式,包括厚数据分析的流程、分析结果的故事化手段^[3]。除学术研究成果外,2017年3月由浙江大学主办的“大数据与厚数据发展前瞻”高端研讨会成功举行^[16]。由此可见,情报学界对厚数据研究热度逐渐升温,厚数据研究也成为大数据时代

背景下与大数据成对出现的重要研究对象。

3.2 厚数据在教育领域的应用

教育大数据、智慧教育是随着信息技术的发展而衍生出来的教育学研究方向,厚数据在教育发展过程中的作用日趋明显。赵兴龙(2017)指出在智慧教育体系构建研究中,应将大数据宽度和厚数据深度结合起来,并强调厚数据的宽度是教育领域应用过程中的重点考量因素^[8]。Sølvi Mausethagen 等(2018)基于对挪威中学教师在“结果会议”中使用知识来源时发现,数据本身虽然很“薄”,但在使用实践时却很“厚”,指出薄数据和厚数据的关系^[17]。涂涛等(2019)通过对大数据与厚数据的关系辨析,提出基于教育大数据的厚数据挖掘^[18]。吴忭等(2019)对美国 AERA2019 年会进行述评,并基于智慧教育研究提出未来应该采用混合的研究方法,回归跨学科研究传统,突破“数据至上”的藩篱^[19]。

3.3 厚数据在公共管理领域的应用

数据成为社会发展的重要生产要素之一,被广泛应用于社会发展的方方面面,厚数据在公共管理领域也开始发挥出独特的优势。姜泽黎等(2017)将大数据与厚数据进行多维度比较,并结合我国海关风险管理的不足,将厚数据思维与海关风险管理工作结合,对我国海关“智慧风险管理”进行深入研究,提出建立海关风险管理工作回路、立体海关风险情报信息系统和海关风险管理厚数据模型^[20]。Annelien Smets 等(2018)提出“公民工具箱”这一试用于研究人员的工具,通过弥合大数据和厚数据之间的差距来应对大数据带来的挑战,使研究人员能够在上下文中积累研究经验,并指出了情景化、语义、分析和人为解释四种策略^[21]。Lu Weisheng 等(2018)通过从政府机构获得的定量大数据与从案例研究和访谈中获得的定性厚数据进行三角剖分,以分析绿色建筑对减少建筑废料的影响^[22]。上述成果从公共管理的不同方面进行研究,并提出了结合厚数据的具体应用方法,体现出厚数据在不同领域的适用性。

3.4 厚数据在医疗健康领域的应用

在大数据时代,医疗大数据已经成为医学与大数据融合研究的热点领域,对比之下,“医疗厚数据”相关研究较少,目前将厚数据应用于医疗健康领域的研究成果主要集中在国外。Kate Winskell 等(2018)以艾滋病纵向叙述数据为基础,进行跨国和纵向研究而开发出数据缩减和管理策略,提出三种数据管理方法:叙事量化特征、叙述摘要及关键词、

专题代码^[23]。Ahn Jin - Ho 等(2020)将厚数据分析方法与丧葬服务业结合,详细收集和分析了当地居民和旅居者的人种志和用户体验数据,指出韩国丧葬服务业的未来发展方向^[24]。Fiadhi Jinan 等(2021)基于虚拟护理研究,提出了一种 VH - CPS 生态系统解决方案,并在该系统中集成了一个可以使护理团队借鉴厚数据分析范式中定性技术的组件,协助推动更深入的分析,使更多的患者处于低风险的护理环境中,并以更低的成本实现更好的结果^[25]。

3.5 厚数据在市场营销领域的应用

国内外有关厚数据在企业创新发展的研究中,较多体现在与企业发展有关的市场营销、财务管理等多个方面。2004年,丹麦知名玩具企业乐高(LEGO)公司在产品市场匹配失败面临破产时,该公司新任首席 CEO Jorgen VigKnudstorp 深入研究儿童玩具需求并收集了大量数据,利用厚数据分析而挽回颓势。2008年,丹麦康乐宝(Colo plast)公司也通过收集、处理并分析有价值的用户厚数据解决了市场份额流失困境。厚数据在商业实践中的成功应用也让学者们开始关注到厚数据所拥有的独特优势和隐性价值。

Michael Alles 等(2014)强调厚数据可以揭示大数据可视化和分析背后的含义,并将厚数据分析方法应用到审计实践中,指出厚数据可以将上下文添加到大数据定量分析中^[26]。叶丽雅(2015)通过多个市场营销案例,指出大数据的优势与风险,强调数据分析应该多一个维度,即厚数据^[7]。Reday 等(2016)关注到小企业销售人员利用信息革命中大数据分析的能力欠缺,以此提出小企业销售人员可以利用人种志、个人访谈等定性研究方法收集厚数据的建议^[27]。银昕(2016)提及两位市场专家指出企业要真正预测未来市场和客户反映,除大数据分析外,还需要厚数据探究用户行为背后的动机,并强调市场营销仍需人类学家^[28]。贾男等(2017)指出我国企业信用体系现有的问题,并提出将厚数据抽样调查应用于体系建设中,展望大数据与厚数据的融合应用^[9]。Thompson(2019)讨论了大数据神话和厚数据机会主义在市场营销中的关系,提出厚数据不能挑战大数据神话,并认为以文化为导向的营销研究人员应该提倡不同的本体论框架^[29]。谭青等(2019)基于生存因素理论研究财务危机的影响因素,提出基于大数据和厚数据的财务危机预警体系^[30]。张晨旭等(2020)通过面向城市合租人群的

冰箱设计策略研究,提出大数据与厚数据融合的基本框架,并以哥本哈根市自行车评估研究案例说明大数据与厚数据研究方法的必要性和实用性^[31]。

上述研究成果从不同研究视角呈现出大数据与厚数据十分紧密的联系,尤其强调了要将大数据与厚数据进行融合,真正体现了“如果没有‘厚数据’,大数据就没有意义”的核心要义^[32]。

3.6 厚数据在社区改造中的应用

在社区改造中需要通过深入探索,提出更加人性化的建设方案与策略,厚数据及相关方法正好贴合了这样的需求。张希煜等(2018)基于厚数据提出了社区更新策略,并提出社区侦探概念,让管理者深入社区,通过数据认知居民需求,用数据实现共情^[33]。张希煜等(2018)基于“厚数据+大数据”对背景鸭子桥社区进行诊断和设计,探索精细尺度公共空间—公共生活评估方法,以激活老旧社区公共生活,进一步验证了“厚数据+大数据”在社区公共服务建设中的重要性和有效性^[34]。郑儒雅等(2020)利用厚数据用户需求调查优势,采用渐进式调查方法对上海松江人乐花园进行深入调研,以迭代分析形式对行为需求数据进行挖掘与重构,提出厚数据支持下的老旧社区公园更新路径^[35]。而霍海鹰等(2020)则基于厚数据理念将 PSPL (Public Space & Public Life) 调研法与定量空间句法创新融合,以此对邯郸市罗城头老旧社区中居民行为、公共生活进行记录和解码,最终以定性与定量结合的分析结果提出老旧改造建议^[36]。

4 厚数据分析方法研究

厚数据分析方法是厚数据研究的重要方面,是厚数据广泛应用于众多场景中得以实现的保障。目前,国内外的厚数据分析方法研究主要集中在人类学方法、数据故事化方法、可视化方法和情感分析方法四个方面。

4.1 人类学方法

厚数据的理论基础来源于人类学,因此,在厚数据应用中人类学方法占据重要地位。吴忭等(2019)在访谈国际知名学习科学专家科学专家戴维·谢弗时,咨询其有关大数据时代背景下学习和测评的观点和经验,戴维教授指出目前数据挖掘风潮下存在一个误区,即缺少用于描述行为行为背后含义的理论模型^[37]。随后,吴忭等(2021)又在数据驱动视角下提出基于量化民族志的教育研究方法,指出量化民族志注重构建“境脉中的理论”和实现“理论饱和”,并提供了半自动编码软件 nCoder 和认知网络

分析两种研究工具,以确保建模结果的可靠性和可解释性^[38]。可见,尽快在大数据时代展开定量分析是效率、速度的体现,但厚数据分析则更注重数据的厚度,更侧重挖掘数据背后的人文精神和因果关系。数据不仅仅只是数字、符号的堆砌,它记载了人类的历史,是人类活动的反映。因此,人类学研究方法也成为数据科学研究领域的一个重要方向。

4.2 数据故事化方法

厚数据分析需要将数据背后的因果关系、故事及意义展示出来。国内学者孙智中等(2020)对比大数据分析后提出了厚数据分析的研究路径,他提到了一个厚数据分析的核心方法,即厚数据结果的故事化,这是一种厚数据定性分析的具体方法^[3]。数据故事化是其重要方法之一,通过数据故事化可以让人们更加了解数据背后的意义、因果关系和故事。中国人民大学朝乐门教授研究团队是目前国内研究数据故事化的核心团队之一。朝乐门等(2019)在《中国图书馆学报》发表《数据故事化:从数据感知到数据认知》一文,提出数据呈现的两种主要途径即数据可视化和数据故事化,并对数据故事化的定义、理论基础、方法工具及应用场景进行了深入研究^[39]。张晨(2021)根据数据故事化研究现状构建了基于用户交互行为分析的数据故事模型^[40],该团队又在《情报资料工作》2021年第2期上发表了“数据故事化:一个值得研究的新方向”的研究专题,主要包括与数据故事化相关的研究成果,该团队对数据故事的自动生成、数据故事叙述的关键技术、数据故事化评价以及可视故事化的特征、方法与应用进行了全面深入的探究^[41-44],其研究基础夯实,研究成果质量高,可以成为厚数据方法研究中的重要参考来源。Thomas Neifer 等(2020)指出数据故事化方法在公司战略挑战和数据分析结果的有效传达上起着决定性的作用^[45]。可以看出,数据故事化是目前厚数据分析中的重要方法之一,也成为数据科学研究的一个新方向。

除此之外,朱梦泽等(2019)对将数据及信息以故事叙述方式生动呈现的叙事式可视化研究进行了综述^[46]。在具体应用中,数据新闻使用数据并以最佳方式叙述故事,且是开放的、可访问的和启发性的^[47],数据故事化作为数字新闻的一项特色也正在迅速得到重视并广泛应用^[48]。杨晓军(2016)认为数据的复杂性决定了新闻更应注重故事化叙事,且受阅读方式、数据呈现与新闻叙述的三重驱动^[49]。档案是人类文化发展中宝贵的历史记录,牛力等

(2021) 基于数字人文环境下档案研究方法的发展,提出了以“发现”“重构”和“故事化”三原语为主线的档案研究路径基本模式,形成了数字人文视角下档案研究的方法论体系^[50]。

4.3 可视化方法

台湾学者宋世祥(2020)在其著作《百工里的人类学家 2:厚数据的创新课》中提出在利用人类学研究方法如民族志、田野调查后,应该促进数据价值的有效传递,并基于厚数据的文化观将其进行创新^[51]。宋教授强调应善用厚数据,化解沟通障碍,因此,他提出了厚数据可视化的四种路径,包括厚数据如何被理解:以同理心贴近对方的语言和经验;如何被看见:以视觉化图表和照片重构场景与脉络;如何被体验:激发五感,搜集具有象征意义的物件;共创历程:邀请厚数据提供者加入共创设计。厚数据研究是跨学科的,为加深数据的厚度、立体度,实现数据的有效传递和沟通,宋教授设计了三大类共六种厚数据可视化工具。第一类是厚数据思考视角工具:使用者基本维度图、元问题图;第二类是脉络重构工具:身体隐喻提问图、使用经验脉络重建图;第三类是需求探索与创新点挖掘工具:换位思考图、人物志+冰山模型图。此外,易旒等(2021)阐述了可视故事化的基本流程,并总结了三种可视故事化方法,包括视觉结构、突出显示和过渡指导^[44],并提出人机交互、数据可视化、文本处理和信息融合四种技术。

4.4 情感分析方法

根据国内外学者对厚数据的相关研究可以看出,在大数据时代不仅要利用好大数据,更要揭示数据的意义和情感,让数据有温度。目前,国内外许多学者利用不同的工具、方法,基于信息技术的量化手段,研究如何有效地揭示数据意义及情感。

Thelwall Mike 等(2009)将数据挖掘和内容分析相结合,探讨了 My Space 评论中情感程度与年龄、性别之间的关系^[52]。何跃等(2018)利用朴素贝叶斯分类器对 2015 年北京新浪微博用户发布的有关“雾霾”的内容进行情感分析,以此分析雾霾舆情、引导公众情绪^[53]。罗玉萍等(2018)利用语义分析技术对学生评教信息进行情感挖掘,建立起情感词库并设置权重,再以 3GWS 工具对学生评教信息中对情感的褒贬程度进行计算,形成学生对教师的评价意见,罗玉萍等以烟台大学相关数据验证了该方法的科学有效^[54]。Kauffmann 等(2019)提出了一个使用自然语言处理技术(NLP)的通用框架,包括情

感分析、文本数据挖掘和聚类技术,以获得针对不同产品特征的基于消费者情感的新评分,并得出结论:企业可以根据客户的意见对客户进行分类来改进推荐系统^[55]。邵珊珊等(2019)利用 Python 和深度学习方法挖掘中美知名电商的评价话语并进行情感分析,旨在促进不同语种间的大数据情感分析研究^[56]。史达等(2020)运用 Word2Vec、LDA 和机器学习(LR、NB、SVM)等方法对 Trip Advisor 的酒店评论数据进行量化处理和情感分析,该研究更加准确地揭示了顾客体验的真实情感,但除了技术支撑外,基于数据挖掘的情感分析也要重视人的因素^[57]。杜洪涛等(2016)在舆情研判研究中指出大数据分析和计算机技术的重要作用,强调在实践中应重视人的主导作用 and 技术的互动耦合^[58]。

从上述研究成果可以看出,在情感分析研究中较常利用的工具包括但不限于 Python、3GWS、Word2Vec、LDA 等,使用的方法包括但不限于深度学习、朴素贝叶斯、内容分析法等。值得注意的是,部分研究成果在阐述数据挖掘相关技术特性时,仍强调了人在数据情感分析时的主导作用,在之后的厚数据研究中应加强“以人为本”的研究思维。尽管上述研究成果中并未直接提及“厚数据”一词,但其对数据背后情感和因果的揭示与“厚数据”内涵较为相近,本文通过梳理部分国内外基于数据挖掘的情感分析方法,旨在为未来厚数据分析方法研究提供参考。

5 厚数据研究展望

尽管厚数据的理论基础可以追溯到 20 世纪 90 年代甚至更久远的人类学哲学思想,但国内外厚数据的相关研究却集中在 2013 年大数据被广泛提及后,主要表现为厚数据与大数据成对出现。由此看来,正是有了大数据才让人们关注到厚数据并对其进行探索。对于这样一个大数据时代的全新概念,未来的研究重点和方向或将集中在厚数据理论研究、厚数据融合研究、厚数据方法研究、厚数据应用研究等多个方面。

5.1 厚数据理论研究

目前,业界和学术界对厚数据都有较为广泛的实践应用和学术研究,但是对厚数据的定义及内涵仍未达成一致,有关厚数据的定义研究仍有待深入。实践需要理论的支撑,理论的产生发展都有其不同的社会背景和领域特色,尤其是在大数据背景下,厚数据作为一个全新的概念,成为数据科学研究中不可忽略的一个分支。因此,未来需要持续不断地对

厚数据定义与理论进行研究,拓展厚数据理论以支撑厚数据实践和应用的能力,全面把握厚数据科学理论对实践的指导作用。

5.2 厚数据融合研究

人类一直都以大数据的海量、高价值为傲,并以此作为根本遵循去进行数据应用,但在技术快速发展的洪流中,人类是否应该停下脚步进行思考,或者回归到传统的分析方法上。当然,最好的发展,就是传统方法和新技术方法的不断磨合、融合,共同进步、发展。大数据与厚数据的结合正是定量分析和定性分析方法的巧妙融合,既能利用大数据的宽度来揭示数据的相关关系,又能利用厚数据的深度来揭示数据背后的因果关系和意义价值。在大数据时代背景下,厚数据的跨学科发展值得关注,学科融合应该更加多元,如厚数据可以与情报学领域中的计量学进行跨学科融合,形成厚数据计量学,利用计量学方法对厚数据隐性知识进行深入挖掘并揭示意义。

5.3 厚数据方法研究

除上文提到的人类学研究方法、数据故事化方法、可视化方法和情感分析方法外,Ethnography Matters 刊载的题为《(大)数据时代的五种混合研究方法》的文章强调“大数据传递数字;厚数据可以传递故事。大数据依赖于机器学习;厚数据依赖于人类的学习”,并提出实时接受受众反馈的民主反思法、自拍工作坊、跟踪采访、移动应用的演练方法和微量人种志五种混合创新研究方法,在这五种混合研究方法中,研究人员概述了他们如何使用创新的方法来充分利用数字数据,这些方法揭示了数据的含义、上下文和故事^[59]。数据的产生归根结底来源于人类活动,在频繁的人类活动和海量数据堆积下,人类社会成为一个偌大的数据库,且人类具备大数据分析中机器所没有或不能达到的揭示情感和意义的的能力,因此“以人为本”是厚数据分析不可或缺的研究思维,将厚数据与大数据分析方法融合创新成为未来揭示数据意义和情感的重要方向之一。

5.4 厚数据应用研究

国内外诸多学者都更倾向于将大数据与厚数据结合在一起进行多场景应用,这既适应了大数据时代发展,又融合了定性分析优势。在大数据和厚数据如何在企业组织中发挥协同效应的思考中,Wang T提出了大数据和厚数据融合发展的多个应用场景,包括但不限于健康医疗、社会网络分析、产品/服务设计和品牌战略等^[4]。基于国内外研究现状可

以发现,厚数据研究场景十分广泛,包括情报学研究、教育大数据、公共管理、人类健康、财务审计、市场营销、社区改造等。随着信息技术、数字人文等多领域的创新发展,厚数据应用场景将会更加多元化。

6 结语

本文首先对当前国内外厚数据的相关定义研究进行综述,可以看出,目前各界对厚数据的定义尚未达成权威一致的共识,但基本形成了以人类学哲学思想为理论基础、以定性分析为核心的内涵。其次,对国内外厚数据在具体场景中的应用进行综述,总结归纳出包括情报学、教育、公共管理、医疗健康、市场营销和社区改造六大应用场景。再次,对国内外厚数据分析方法进行认真思考和剖析,归纳出目前厚数据分析研究中的四大类研究方法,主要包括人类学方法、数据故事化方法、可视化方法和情感分析方法。最后,基于国内外厚数据研究的最新进展和发展趋势,指出厚数据未来的研究重点和方向,可围绕厚数据理论研究、厚数据融合研究、厚数据方法研究和厚数据应用研究四个方面进行深入探究。

参考文献

- [1] 文庭孝,姜珂妍,赵阳,等.大数据时代的信息分析变革研究[J].图书情报知识,2015(5):66-73.
- [2] 高兆明.“数据主义”的人文批判[J].江苏社会科学,2018(4):162-170.
- [3] 孙智中,张晨.基于厚数据的信息分析:内涵与模式[J].情报资料工作,2020(3):69-75.
- [4] Wang T. Big Data Needs Thick Data [EB/OL]. [2021-04-20]. <http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/>.
- [5] Madsbjerg C. The Power of ‘Thick’ Data [EB/OL]. [2021-04-20]. <https://www.wsj.com/articles/the-power-of-thick-data-1395443491>.
- [6] 刘琳,张曙霞.央视大调查的数据“秘密”[EB/OL]. [2021-04-26]. <https://mp.weixin.qq.com/s/6PVH13O1gEhCnCyc2QQvGg>.
- [7] 叶丽雅.厚数据,给大数据以灵魂[J].IT经理世界,2015(6):54-55.
- [8] 赵兴龙.核心素养视角下的智慧教育体系构建[J].现代远程教育研究,2017(3):34-43.
- [9] 贾男,刘国顺.大数据时代下的企业信用体系建设方案[J].经济纵横,2017(2):40-44.
- [10] 克里斯蒂安·马兹比尔格.意会:算法时代的人文力量[M].中信出版社,2020:63-82.
- [11] Ang Y Y. Integrating big data and thick data to transform public services delivery [R]. IBM Center for The Business of Government,2019:1-46.
- [12] 武涛,刘叶婷.“进化”中的大数据:新特性、新变化、新态势[J].信息系统工程,2015(3):101-103.
- [13] 袁曦临.制约我国智库研究与发展的瓶颈问题——跨学科研

- 究与专题性研究资源保障[J]. 情报资料工作, 2017(5): 99-104.
- [14] Bornakke T, Due B L. Big - thick blending: a method for mixing analytical insights from big and thick data sources [J]. Big Data&Society, 2018(1): 1-16.
- [15] Fiaidhi J, Mohammed S. Thick data: a new qualitative analytics for identifying customer insights [J]. IT Professional, 2019(3): 4-13.
- [16] “大数据与厚数据发展前瞻”高端研讨会在浙江大学成功举办[J]. 浙江大学学报(人文社会科学版), 2017(3): 92.
- [17] Sølvi Mausethagen, Tine Prøitz, Guri Skedsmo. Teachers' use of knowledge sources in 'result meetings': thin data and thick data use[J]. Teachers and Teaching, 2018(24): 37-49.
- [18] 涂涛, 胡柯铭. 一极两仪: 教育大数据与厚数据关系辨析[J]. 中国电化教育, 2019(8): 18-22.
- [19] 吴怵, 彭晓玲, 胡艺龄. 教育研究的去芜存菁之路: 从多模态叙事到证据公平——美国 AERA2019 年会述评[J]. 远程教育杂志, 2019(4): 13-23.
- [20] 姜泽黎, 沈斌. 浅议“厚数据”思维下的我国海关“智慧风险管理”[J]. 中国市场, 2017(30): 154-156.
- [21] Annelien Smets, Bram Lievens. Human sensemaking in the smart city: a research approach merging big and thick data[J]. Ethnographic Praxis in Industry Conference Proceedings. 2018(1): 179-194.
- [22] Lu Weisheng, Chen Xi, Peng Yi, et al. The effects of greenbuilding on construction waste minimization: triangulating 'big data' with 'thick data' [J]. Waste management. 2018(79): 142-152.
- [23] Kate Winskell, Robyn Singleton, Gaele Sabben. Enabling analysis of big, thick, long, and wide data: data management for the analysis of a large longitudinally cross-national narrative data set[J]. Qualitative Health Research. 2018(28): 1629-1639.
- [24] Ahn Jin - Ho, Lee JeungSun. A study on the direction of funeral service focused on thick data analysis[J]. Journal of Service Research and Studies. 2020(10): 85-96.
- [25] Fiaidhi Jinan, Mohammed Sabah. Virtual care for cyber - physical systems (VH - CPS): NODE - RED, community of practice and thick data analytics ecosystem [J]. Computer Communications. 2021(170): 84-94.
- [26] Michael Alles, Miklos A, Vasarhelyi. Thick data: adding context to big data to enhance auditability[J]. International Journal of Auditing Technology. 2014(2): 95-108.
- [27] Reday P, Ulusoy E, Kos A, et al. Big data versus little and thick data and the manufacturer's representative[J]. International Journal of Business Research, 2016(4): 19-28.
- [28] 银昕. 大数据营销如何突破传统 CRM? [J]. 商学院, 2016(4): 74-75.
- [29] Thompson. The 'big data' myth and the pitfalls of 'thick data' opportunism: on the need for a different ontology of markets and consumption [J]. Journal of Marketing Management. 2019, (35): 207-230.
- [30] 谭青, 许佳丽. 基于生存因素理论的财务危机影响因素研究[J]. 天津商业大学学报, 2019(4): 14-20.
- [31] 张晨旭, 张凌浩. 厚数据背景下面向城市合租人群的冰箱设计策略[J]. 大众文艺, 2020(2): 102-103.
- [32] 刘源隆, 洪治. 十年“大调查”, 一本中国账[J]. 小康, 2016(5): 55-57.
- [33] 张希煜, 欧阳亦琛. 社区侦探: 基于厚数据的社区更新策略[J]. 建筑创作, 2018(5): 78-95.
- [34] 张希煜, 茅儒睿, 邢晓旭, 等. “厚数据+大数据”激活老旧社区公共生活——以北京鸭子桥社区为例[J]. 北京规划建设, 2018(5): 4-10.
- [35] 郑儒雅, 汤晓敏, 王云. 厚数据支持下的城市老旧社区公园更新路径研究——以上海松江人乐花园为例[J]. 安徽农业大学学报, 2020(6): 986-995.
- [36] 霍海鹰, 王士坤, 李佳, 等. “厚数据”视角下邯郸罗城头老旧小区公共空间改造研究[J]. 城市建筑, 2020(35): 12-15.
- [37] 吴怵, 胡艺龄, 赵玥颖. 如何使用数据: 回归基于理解的深度学习和测评——访国际知名学习科学专家戴维·谢弗[J]. 开放教育研究, 2019(1): 4-12.
- [38] 吴怵, 彭晓玲. 量化民族志: 一种融合定性性与定量的教育研究方法[J]. 现代远程教育研究, 2021(2): 63-72.
- [39] 朝乐门, 张晨. 数据故事化: 从数据感知到数据认知[J]. 中国图书馆学报, 2019(5): 61-78.
- [40] 张晨. 基于用户交互行为分析的数据故事建模方法研究[J]. 情报理论与实践, 2021(2): 185-191.
- [41] 朝乐门. 数据故事的自动生成与工程化研发[J]. 情报资料工作, 2021(2): 53-62.
- [42] 张晨, 朝乐门, 孙智中. 数据故事叙述的关键技术研究[J]. 情报资料工作, 2021(2): 73-80.
- [43] 孙智中, 朝乐门, 王锐. 数据故事化的评价与改进[J]. 情报资料工作, 2021(2): 81-89.
- [44] 易旒, 朝乐门, 张晨. 可视故事化: 特征、方法与应用[J]. 情报资料工作, 2021(2): 63-72.
- [45] Thomas Neifer, Dennis Lawo, Paul Bossauer, et al. Data storytelling als kritischer erfolgskfaktor von data science [J]. HMD Praxis Der Wirtschaftsinformatik. 2020(57): 1-14.
- [46] 朱梦泽, 赵海英. 叙事式可视化综述[J]. 计算机辅助设计与图形学学报, 2019(10): 1719-1727.
- [47] Rogers S. Data journalism matters more now than ever before [EB/OL]. [2021-04-26]. <https://simonrogers.net/2016/03/07/data-journalism-matters-more-now-than-ever-before/>.
- [48] Ojo A, Heravi B. Patterns in award winning data storytelling [J]. Digital Journalism, 2018(6): 693-718.
- [49] 杨晓军. 数据新闻故事化叙事的可能性及思维路径[J]. 编辑学刊, 2016(1): 114-118.
- [50] 牛力, 高晨翔, 张宇锋, 等. 发现、重构与故事化: 数字人文视角下档案研究的路径与方法[J]. 中国图书馆学报, 2021(1): 88-107.
- [51] 宋世祥. 百工里的人类学家 2: 厚数据的创新课 [M]. 台北: 果立文化出版. 2020: 218-286.
- [52] Thelwall Mike, Wilkinson David, Uppal Sukhvinder. Data mining emotion in social network communication: gender differences in MySpace [J]. Journal of the American Society for Information Science & Technology. 2009(61): 190-199. (下转第 28 页)

Integrated Development of "Rural Library + ": Research on the Origin, Mode and Practical Path

Wang Xiongqing

(Xiangtan University library, Xiangtan, Hunan 41105, China)

Abstract With the continuous expansion of the connotation of the national reading promotion service in China, the construction and popularization of rural libraries have played a key role in the strategy of rural revitalization, promoted the farmers' cultural competence, and improved their mental outlook. During the period covered by the 14th Five - Year Plan, poverty alleviation should combine ambition alleviation and intelligence alleviation, transform the cultural poverty alleviation from a blood transfusion mode to a blood producing mode, and stimulate the integrated development of rural revitalization and the innovative mode of "rural Library", networking business and cultural poverty alleviation.

Keywords "Rural Library +" Project. Rural revitalization. Reading promotion. Innovation. Cultural poverty alleviation

(上接第 14 页)

[53] 何跃, 朱婷婷. 基于微博情感分析和社会网络分析的雾霾舆情研究[J]. 情报科学, 2018(7): 91-97.

[54] 罗玉萍, 潘庆先, 刘丽娜, 等. 基于情感挖掘的学生评教系统设计及其应用[J]. 中国电化教育, 2018(4): 91-95.

[55] Kauffmann E, Peral J, Gil D, et al. Managing marketing decision-making with sentiment analysis: an evaluation of the main product features using text data mining [J]. Sustainability, 2019 (15): 1-19.

[56] 邵珊珊, 王立非. 基于语言大数据挖掘的电商英汉评价话语情感分析[J]. 外语电化教学, 2019(5): 76-84.

[57] 史达, 王乐乐, 衣博文. 在线评论有用性的深度数据挖掘——基于 TripAdvisor 的酒店评论数据[J]. 南开管理评论, 2020(5): 64-75.

[58] 杜洪涛, 孟庆国, 王君泽. 互联网数据在舆情研判中的效用研究[J]. 中国软科学, 2016(4): 34-44.

[59] Ethnography Matters. Five Mixed Methods For Research in the (Big) Data Age [EB/OL]. [2021-04-26]. <http://ethnographymatters.net/blog/2016/05/03/five-mixed-methods-for-research-in-the-big-data-age/>.

[作者简介] 李成熙, 中南大学湘雅医学院医药信息系 2017 级硕士研究生; 文庭孝, 中南大学湘雅医学院医药信息系教授。

[收稿日期] 2021-07-20

(肖蔚 编发)

Thick Data: A Research Review

Li Chengxi Wen Tingxiao

(Central South University, ChangSha, Hunan 410013, China)

Abstract By reviewing the theoretical origin and the development of thick data and using WoS and CNKI to retrieve relevant literature, this paper sorts out the current research status of thick data both at home and abroad, analyzes and interprets the research hotspots and frontiers, and provides support and suggestions for future research. Existing research on thick data mainly focuses on the definition, the application scenarios, and the analysis methods of thick data. Future research should further explore basic research issues, and continue investigating the integration, methods and application of thick data.

Keywords Thick data. Big data. Thick description. Qualitative analysis. Data storytelling. Visualization.