

大数据驱动社会科学研究的实践向度〔*〕

○ 陈 潭¹, 刘 成²

(1. 广州大学 公共管理学院, 广东 广州 510006;
2. 北京师范大学 政府管理学院, 北京 100875)

〔摘 要〕随着大数据在市场范畴的广泛应用, 社会运行数据、网络交互数据、人类行为数据等数量庞大且结构多样的数据集合在重新定义传统的市场运行逻辑的同时, 也必将大幅扩展社会科学研究中原始数据的维度、量度、深度和精度, 多方位、多角度、动态地描述、解释人类行为和社会现象, 更为精准地探寻社会运行规律。基于此, 在大数据时代, 社会科学研究将凸显出数据驱动社会科学研究的特征, 传统研究方法及其研究范式将被重塑和再造, 大数据时代社会科学研究的实践向度将逐渐从经验研究转向真实研究、个案研究转向全面研究、断面研究转向交叉研究、表面研究转向深度研究。

〔关键词〕大数据; 社会科学; 实践面向

DOI: 10.3969/j.issn.1002-1698.2017.07.013

一、大数据引领社会科学研究范式的变革

随着互联网技术的不断发展, 人类社会逐渐步入大数据时代, 数据资源正和土地、劳动力、资本等生产要素一样, 成为促进经济增长和社会发展的重要资源。任何一个行业和领域都会产生有价值的信息, 而对这些信息的统计、分析、挖掘则会创造出意想不到的价值和财富。

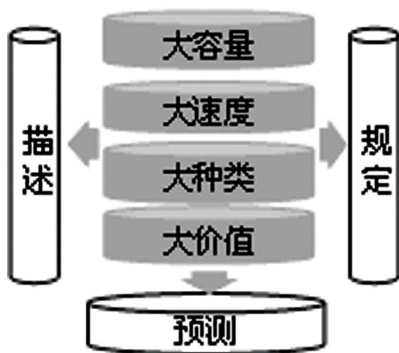
根据国内外学者的深入研究, 大数据具有四大基本特征, 即 4V 特征: (1)

作者简介: 陈潭(1969—), 法学博士, 广州大学公共管理学院教授, 主要从事大数据政务和创新治理研究; 刘成(1990—), 北京师范大学政府管理学院博士研究生, 主要从事创新治理研究。

〔*〕本文系国家自然科学基金 2015 年重点项目“国家大数据战略实施的关键路径及其行动策略研究”(15AZD077)阶段性成果。

Volume 是指数据容量特别大,并呈现出指数级增长态势;(2)Velocity 是指数据增长速度快,包括数据更新换代速度和数据运算速度;(3)Variety 是指数据种类繁多,涵盖各种来源、形式和结构的数据;(4)Value 是指价值性,通过数据挖掘服务实践应用。同时,大数据也具有三大基本功能:(1)Descriptive(描述)是通过对数据的统计分析,描述数据表现出的现象与规律;(2)Prescriptive(规定)是利用历史数据建立分析模型和规范化的分析流程,实现对连续数据流的实时分析;(3)Predictive(预测)是通过对数据的深层挖掘构建预测模型,实现对未来发展趋势的预测^[1]。

图1 大数据特征与功能



由此可知,通过挖掘海量数据能够让潜藏的社会规律可视化、机械流程智能化、经验决策科学化,能够有效提升管理的实时性、准确性和科学性。随着大数据在制造业、商业、服务业的“大显身手”,它也赢得了社会实践领域特别是市场经济领域的肯定和称赞。那么大数据又将如何驱动社会科学研究的转向呢?

社会科学自产生以来,历经了几个世纪漫长而又复杂的演进和蜕变,其研究内容、主题、范域也愈加细化。同时,随着研究理念的更新、研究方法的换代,社会科学的研究范式也发生了不同程度的迭代和革新。吉姆·格雷(Jim Gray)提炼并概括了社会科学研究范式的四次演变,他认为,社会科学的发展经历了经验型、模型和概念型及复杂现象模拟型三次研究范式,即将到来的第四次范式革新,被称作大数据时代的数据驱动型研究范式。文艺复兴前后,自然科学发展得如火如荼,极大地促进和助推了社会科学的发展,也为社会科学提供了现成的效仿模版。在自然科学的影响下,社会科学接受了自然科学中的假设与验证的研究范式,坚定地认为人类社会存在一种普遍的、永恒的本质规律,并初步建立了以实验为主要方式的经验主义范式。进入前计算机时代之后,受到生物学等学科的影响,社会科学建立了以有机整体论的社会观为前提,通过结合理论构建和实证研究,以整体的、系统的视角探索社会运行的规律^[2]。而在计算机时代,人类通过数字运算而获得归纳和推理能力得到大幅提高,通过利用相似原理建立研究对象的模拟,通过计算模型间接地研究原型的规律性,因而计算机带来的复杂现象模拟技术被广泛地运用到国民经济、企业战略管理、国家治理等社会

科学领域。

表 1 社会科学研究的四次范式

范式	性质	形式	时间
第一	实验科学	经验主义	文艺复兴前后
第二	理论科学	模型和概念化	前计算机时代
第三	计算科学	复杂现象模拟	前大数据时代
第四	探索科学	数据驱动、数据挖掘	大数据时代

随着社会生活与科学技术的深度结合,社交媒体已成为当代社会交互的主要媒介,深刻地变革着人类传统的生产方式、生活方式、思维方式、交际方式、行为方式,人类交互和行为的可视化和平台化不仅加速了人类认识和改造自然世界的进程,也标志着人类认识和改造自身世界进入到更为客观、精准、全面的历史性阶段。加之,大数据以及数据分析工具和挖掘技术的蓬勃发展,社会科学研究开始步入了以大数据为支撑的第四次研究范式,海量、即时、动态、多样的数据供给和先进的数据挖掘技术成为了引领新时代社会科学研究的范式。在数据资源和挖掘技术的支撑下,社会科学研究步入了以数据挖掘为核心方法的研究范式,数据驱动型学术研究将成为社会科学研究的主体。毫无疑问,这种数据驱动型学术研究将引导社会科学从经验研究转向真实研究、表面研究转向深度研究、个案研究到全面研究、断面研究到交叉研究。

二、大数据驱动社会科学走向真实研究

社会科学的研究领域是整个人类社会,是人类社会中无数灵动的个体和他们之间的交互所构成的集合,是一个多元的、非线性相关的、复杂的适应系统,在这个适应系统中,交织着异常复杂的人际互动和价值传播,极具随时性和易变性,是一种动态的存在^[3]。借助大数据带来的数量巨大、种类繁多、实时动态、富含价值功用的多元数据,现代社会科学理应形成以真实的人类社会为数据源的庞大数据集,构建动态实时的数据抓取和分析系统,将数据作为了解和认识人类社会运行的本质规律的基础和核心,注重单个个体的特殊性和异质性以及整体社会的多维性和复杂性,以便充分发挥大数据与社会科学有机结合之后的倍增效益。

由于数据获取方式的限制,大数据时代之前的社会科学研究数据一般属于非原始数据(Made Data),这些数据是由研究者在确定的研究假设下,从人为框定的抽样样本或人群中获取信息和数据,通常数量较小且较为简单,这就导致数据收集的成本增加和周期变长,同时由于数据采集方法的限制,非原始数据难以保证数据是研究对象真实的心理动机或现实境况。例如,传统社会科学使用的问卷调查或访谈常常会遇到一些个人隐私,甚至较为敏感的问题,诸如个人收

人、年龄、婚姻状况、堕胎等问题,这些问题使得受访者不愿意或隐藏真实想法,也容易导致所采集的数据变成毫无意义的“虚假”数据。但在大数据时代,数据驱动的社会科学研究的数据是原始数据(Found Data),是客观存在而非专门服务于特定研究目标的数据,也被称之为“元数据”,具有数量庞大、复杂多样且角度多维的特点。原始数据以终端研究对象为数据获取源头,描绘复杂的人类感官世界,展现个体真实的内心世界,体察驿动的心理动态,挖掘丰富的人际交互,探索人类社会总体价值走向,能够有力支撑流行病防控、城市规划、公共政策、公共安全等社会科学领域的研究。例如,通过深度分析和挖掘社交网络中海量的交流记录,对在线交流数据的分层处理能够实现政府决策注意力从集中于意见领袖而忽视普通大众到两个群体意见合理平衡的转变。

在社会关系范域上,大数据意味着人际交互可视化,也开创了理解人类复杂意识世界的革命性方式。一项以朋友关系为切面理解社会关系变革的研究就是依托于社交媒体中存留的海量人际交互大数据而开展的全新的研究图景。该研究验证了海德尔的社会平衡理论中的两个主要特征,互惠性和传递性。作者将朋友定义为定向关系,即人物 A 把人物 B 视为朋友,记作(A→B)而非(B→A)。他们还能够预测友谊,不仅能识别朋友和非朋友关系,也能预测非互惠朋友人物 A 与人物 B 的友谊可表示为(A→B)且(B→A)。研究者利用麻省理工学院的现实挖掘工程(MIT Reality Mining project),在用户的手机里安装一个应用软件,直接从现实世界采集用户位置、邻近用户、文本信息和通话记录等用户行为和交互数据。通过利用支持向量机(Support Vector Machine)模型分析和挖掘用户距离、已接电话、未接电话等手机信息中的特征,进而预测研究对象间的友谊^[4]。

较之于自然科学,社会科学的研究范域是整个人类社会,个体的心理和行为活动无时无刻不在发生变化,而由无数个体构成的社会系统则更为复杂多变、难以预测,这也就注定了其复杂性和动态性是自然科学无法企及的,进而也就决定了社会科学研究必须考虑数据即时性问题,但传统研究中的数据有很大的时滞性,许多研究成果问世时可能已经过去多年,研究对象、研究问题都已发生剧烈变化,这给研究的真实性提出了巨大的挑战。这也就要求研究者要想办法实时地反映研究对象的真实境况,只有确保数据的即时性才能保证研究结果的真实性和有效性,也才能体现社会科学研究特别是实证研究的价值。

如今,人类行为活动和交互信息的平台化和可视化已是大势所趋,推特每月的活跃用户超过 2.8 亿人,每天的交互信息超过 5 亿条,脸书每月的活跃用户更是超过 13 亿。为了给用户和第三方机构提供最新的数据信息,这些公司特地提供被称之为“数据流”的服务,能够在特定的时段给指定的用户自动发送通知信息。所以,这些数据具有实时的特点,其内容会随着时间的变动而变化,提炼这些实时的数据有利于挖掘事件背后的规律^[5]。达斯等(Dass, P.J.H. et al)对交通密集度做了实时研究,通过收集散布在街道的交通探测器,他们获取特定路段或路口过往车辆的速度、时长,并且能够将数据更新频率控制在一分钟之内,极

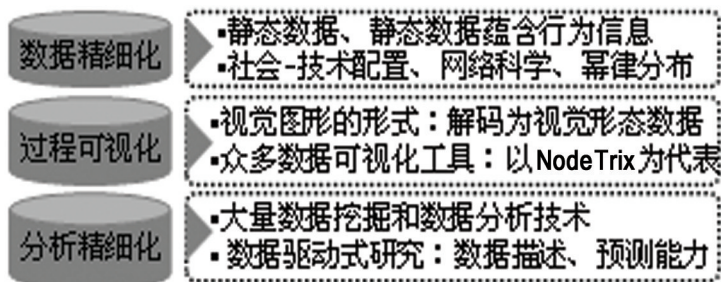
大地缩短了数据抓取的实时性,有效地提升了研究的真实度。

社交媒介平台的流行也有利于自然灾害的防御和治理,身处自然灾害发生地或附近地区的用户能够同步在推特、脸书、微信等社交媒体平台上发布灾害现场的照片、博客及视频,这些信息可以让社会公众及时了解灾害情况,也有助于防灾部门及社会组织获取灾害现场实时的反馈数据。在2010年的海地地震和俄罗斯森林大火以及2012年的飓风“桑迪”等自然灾害发生期间,一些慈善组织建立了基于网页的人工灾害地图,分析和挖掘新闻报道、社会媒体以及政府预警信息,为公众提供实时的灾害地图。但面对海量的包含了结构化和非结构化的异质性数据集,数据的自动化处理成了一大难题,同时数据的真实性也存疑。为了提高数据的自动化处理能力以及数据的可靠性,米德尔顿等学者研发了全新的实时灾害地图平台,一方面该平台利用地名词典、街区地图、自发地理信息(Volunteered geographic information)获取精确的地理位置,同时利用数据分析技术转化各类结构化及非结构化数据,进一步提高了街区层级的区域定位的精确性^[6]。

三、大数据驱动社会科学走向深度研究

进入20世纪以来,定量研究方法一直是社会科学领域发展重要的方法之一,也是社会科学被西方学术界真正视为科学门类的理由和基础。尽管定量研究的兴起弥补了纯粹定性研究在可验证性以及客观性方面的不足,但随着社会关系的越加复杂,社会活动的越加丰富多样,传统定量研究已经无法应对当前社会科学之于数据获取、数据管理以及数据分析层面的新要求和新挑战。相对于大数据,由于数据易获取性的缺陷导致数据的量级以及数据精度都处于下风,而数据分析、数据挖掘能力的局限性也制约着数据利用的深度性和精准性,致使传统社会科学研究较为表面,很难真正达到理想的精致化、细粒化研究。较之于传统定量研究方法,大数据驱动的社会科学则是深度研究,这是因为:一是大数据能够跟踪和记录人类行为和交互活动,实时获取海量结构化及非结构化的可供精细分析的海量数据;二是大数据赋予了社会科学更易于人类接收信息的数据可视化平台;三是大量深度学习工具带来的数据分析的精准化。

图2 大数据驱动社会科学的深度研究



第一,数据的精细化。以数据可被分析的抽象程度为基准,大数据大体可以分为两大类型:一是静态数据,例如与标准算法相结合的信息供给;二是静态数据中蕴含的行为信息,数据价值的体现并非源自这些静态数据本身,而是数据收集的类型以及随后的数据分析抽象程度。在社会—技术配置的支持下,使用计算机的人类之间以及人类与技术设备的交互被完整的存储,例如地理位置、性别、年龄等,更进一步说,每天我们使用网络时都会留下丰富、精准、细致的行为大数据,这些数据并非简单的数字陈列,它是开启人类探究个体、组织、群体、社区的行为方式和运行规律新篇章的全新方法,具有巨大的理论和实践价值。网络科学就是通过抽象的数据分析研究由个体构成的社会关系网络的新兴学科。网络科学的重要成果就是在自组织机制下社会网络会形成具有小世界现象的服从无标度(Scale-free)分布的幂律分布(Power-law distribution)网络;幂律分布(Power-law distribution)的典型特点就是在网络中存在少数但功能很大的具有强连接能力的节点,正是这些具有高度分布取值的节点在社会网络演化中扮演重要的角色^[7]。

第二,过程的可视化。如今,数据获取和收集已经不再是困扰社会科学家的“死穴”,问题的关键在于如何发挥数据的功用,如何挖掘隐藏于数据背后的实际价值,这其中数据本身以及处理后数据的呈现方式是尤其重要,因为数据的呈现方式直接影响信息传递的效果,也决定着信息接受者理解数据的速度和效率。数据可视化被称作数据视觉呈现的科学,通过视觉图形的形式更为清晰、有效地把数据中信息单元的特质和变量展现出来,为人们理解社会事件以及社会事件关系提供了一种全新的方式^[8]。相比单纯的数字型数据形态,被解码为视觉形态的数据会让人们更容易辨别和理解数据间的内在关系和逻辑模式,以简洁、直观的视觉形态呈现的数据、文本等信息可以实现在单位时间内输出更多维、更丰富、更细致的内容。可视化工具在新闻业的应用体现了其独到的价值功用,使得编辑者既能发挥传播信息的功能性也可兼顾它的美学展现,同时也提升读者汲取和分析信息的便利度。Node Trix就是众多数据可视化工具中的一种,在节点连接图解的混合网络代表法的基础上,社区能被各种矩阵所代表,毗连的矩阵可以代表社区与社区间的关系,而正常连接代表社区内部的关系。对于社会网络分析来讲,它的价值功用体现在两方面,一是能够以节点连接图解的方式深度地解构和呈现社会网络结构,二是提高社区细节分析矩阵的易读性^[9]。

第三,分析的精确化。大数据时代的来临让原本隐藏的海量数据得以显现的同时,也催生了大量数据挖掘和数据分析技术的问世,诸如 Node XL, Gephi, DMI Issue Crawler 等一系列深度学习工具和数据分析工具已经在商业领域得到了大量使用。虽然大部分的数据分析工具来自于商业领域,基于学术研究的数据分析工具还仍处于发展的初期,然而,随着人类社会逐渐跨入大数据时代,社会科学家必然要学会如何利用这些数据分析工具。正如 Savage and Burrows 所言,数据分析工具不应该仅仅局限于少数的内行专家或商业人士。在兆级别

甚至千兆级别的数据催生了海量的微观等级的数据信息,以个体人为中心信息收集单位的大数据采集系统不断涌现。为了挖掘隐藏于数据背后的规律,大量的数据分析和挖掘工具不断问世,特别是数据描述和数据预测能力的提升使得社会科学研究逐渐转向为数据驱动式研究。数据分析方法迭代至今,能够描述和解释各种不同类型的流数据,例如视频、文本及 GPS 数据等,得以记忆和留存。通过深度分析和挖掘这些数据,研究者可以掌握每个个体留存在各种社交媒体、网站、数据存储器的数据信息,从而可以分析他们在适用前的使用行为及适用后的使用行为。以消费者行为研究为例,因为有了超个性化的产品和服务的消费模型的数据支持,研究者能够有条件专注于消费行为观察的研究而非局限于单纯的消费动机研究,因为精细的消费者行为感知是非常有可能影响消费者网页浏览和网上购物等行为,是预判消费动机的最为直接和有效方式^[10]。

四、大数据驱动社会科学走向全面研究

作为社会科学研究领域中基本的研究方法之一,个案研究方法具有悠久的历史,被广泛应用于社会科学研究中。它通过“窥一豹而见全身”的方式、以个案为切入点,从而得以探求社会问题的整体境况。然而,对于个案研究,备受学者诟病的是案例选择是否典型,如若不够典型,那么得出的整体性结论便有失偏颇,即便案例具有代表性,由于个案之间的特性的存在,依然无法全面反映整体的状况。有学者认为,个案研究一直面临着特殊性与普遍性、具体与概括之间的矛盾,即便建立了所谓“科学”的抽样系统,内在的代表性和可推论性仍无法保证,研究策略依然倾向于描述和细节,只是对特殊性情有独钟的个案研究难免会陷入“集体失声”的困境^[11]。

随着互联网像细胞一样注入到人类社会的各个方面,虚拟的互联网世界在人类行为过程中扮演的角色也与日俱增,海量人类行为、交流等信息和数据被永久地记录在了社交网络、移动终端等数字虚拟平台,不仅能够记忆、解释和揭示我们的行动轨迹和行为规律,同时也为以人为研究中心的社会科学提供了量级、结构、维度空前的数据保障,也有效地保证了社会科学研究从个案研究走向全数据驱动的全面研究。目前来看,存留在数字虚拟数字平台的数据规模已高达 TB 级别, PB 级的数据库也不少见,以后 EB、ZB 甚至 YB 级的数据库也会陆续出现。这些数据范域从最小的数据集合到最大的数据集合,可以划分为三大层面,分别是微观型数据集合、中观型数据集合以及宏观型数据集合。微观型数据集合是大数据范域中数据级别最小的集合,主要收集和记录个体的人在网页浏览、点击偏好、行为轨迹等个人层面的数据,甚至在人类感情交互的研究中,时常使用纳米和原子级的数据,用来更为精确地评判人类的感情投入和倾向^[12];中观型数据集合是比微观型数据更高级别的数据范域,主要包括社交媒体中人们交互中产生的互动细节以及针对某一话题的态度意见;宏观型数据集合是最高级别的数据集合,包括国家以及跨国层面的国家安全、社会经济、区域消费、教育

医疗以及地理定位等信息,例如不同城市和地区的电力使用情况,国家的犯罪率等信息^[13]。

图3 大数据驱动社会科学的全面研究



微观数据集并不意味着数据的数量微观,而是指对于微小个体的数据采集,也就是说,当海量的微观数据汇集成一个数据集,数据的量级也依然庞大。2010年美国大选期间,Robert M.Bond等研究者就给6100万“脸书”用户随机发放了有关政治运动的信息,研究结果显示这些信息直接影响了数以千万计选民的政治表达、信息获取以及最终的投票行为,不仅如此,这些信息不仅影响信息接受者自身也影响他们的朋友甚至朋友的朋友,感情纽带的深厚程度也直接决定着信息的传播效果,并且信息传播的社会功能要比信息本身更有价值,更为重要^[14]。

中观数据包含一个地区中部分或整体人群的行为,例如在一个城市或社会里追踪特定人群的消费行为及交流数据^[15]。市场和社交媒体场域下的全新数据收集和数据感知方法赋予了研究者在更为整合的市场和社会环境中探测和分析人类行为方式。这类例子不胜枚举,团购、网上购物、交通流量及社会情感,等等。网上招聘网站汇集了一个国家或地区海量的人力资源数据,能够反映该国家或地区的就业率、失业率、市场紧缺工种以及劳动力需求结构的转变。基于市场和社会层面的集体行为研究构成了社会层级的数据集,是政府及第三方机构感知社会运行、提升社会服务需求和质量的有效途径,可以为城市交通堵塞、城市居住质量、公民信用评估等社会问题的解决提供有力的全数据支撑。

宏观数据集通常是跨区域、跨城市甚至跨国家的各类信息,具有高度的整合性和复杂性,是研究跨区域、跨城市、跨国家的综合性社会问题的必要保证。罗曼·蒂莉(Roman Tilly)研究和分析了来自互联网社交媒体的宏观时空旅游评论信息的质量,并从信息的完整度、精确度和实时度加以衡量,共收集了232个国家的789万条旅游相关的交互信息,通过深度数据挖掘,得出如下结论:社交媒体是宏观旅游信息的重要传播源,可以用来评估旅游景区的形象和知名度^[16]。

五、大数据驱动社会科学走向交叉研究

在大数据逐渐渗入学术研究范域的进程中,一些诸如安德森(Anderson)等较为激进的学者甚至断言大数据在社会科学研究中的应用使得传统社会科学理论变得毫无价值,他们认为大数据将完全颠覆现行社会科学研究范式,数据驱动社会科学研究将完全取代现有研究形态。但更多的学者对此嗤之以鼻,在他们看来,大数据技术与社会科学的有机结合并非单独学科、单一路径能够胜任,需要更为广泛的学科范域和途径,是不同学科合作的结晶。由此,大数据驱动的社会科学将是跨学科、跨专业、跨领域的交叉性研究,而非单学科、单专业、单领域的断面研究。未来的社会科学研究将是理论支撑与数据分析相互补充、互相融合的局面:一方面延续传统社会科学在构建理论、分析归因、演绎推测上的传统优势;另一方面发挥和结合其他学科领域特别是数理学科在抓取和处理海量、多样、非结构化数据资源的独特专长。

理论框架伴随数据生成到应用的整个过程,是镶嵌在数据使用背后的内在逻辑,也是研究者运行数据研究社会现象的中枢和纲领,而理论框架的搭建离不开社会科学领域内透彻掌握各专业基本理论的研究者。诚如惠普实验室的物理学家赫伯曼(Huberman)所言:具备良好社会科学专业背景的学者是能否提出有价值研究议题、能否让研究驶入正确轨道的关键。数据量级再庞大、再精细、再多维,数据本身不能得出任何有价值的研究结论,数据分析思维才是中枢神经,拥有制定研究框架的研究者才是操盘手。例如,为了判别和寻找互联网世界中的意见领袖,研究者必须要在浩如烟海的互联网数据和交互信息中确定搜索域;假定研究目标是判定政治抗议中的意见领袖,以政治抗议为核心的数据和信息便是搜索域,这项任务应由政治学专家来完成,因为没有人比他们更擅长政治抗议所包含的关键信息^[17]。

大数据的运行周期(收集、存储、管理、挖掘、预测)是以计算机科学、统计学、数学等数理学科知识为基础,这些知识和技能并非传统的社会科班出身的学者可以胜任^[18]。社会科学家通常没有受过计算机科学和信息科学的专业训练,难以具备收集、获取和处理在线数据的基本计算机技能,也无法利用 API(Application Programming Interface)等在线平台获取目标用户数据。即便社交平台中贮藏着海量有价值的数据,社会科学家也只能望而兴叹,无法真正发挥这些数据的功用。获取数据只是大数据运行中的首要环节,大数据的关键是数据挖掘和预测功能,这才是探求海量数据间的变量相关性以及判定变量因果关系的核心环节。然而,数据分析和挖掘技术要求更为精湛的计算机能力,需要扎实的计算机语言、网络科学、机器学习等信息科学领域的知识。对于擅长理论构建和思辨分析的社会科学家来说,短时间内掌握这些计算机和信息科学领域的专业知识并非易事,因此打破传统学科壁垒,构建以分析方法为核心的跨学科协作模式是未来社会科学研究快速发展的必然趋势。

一部分具有前瞻眼光的社会科学研究者们预知到了大数据与社会科学研究深度融合的必然性,他们立足于不同学科界域,通过嫁接计算机科学、统计学等学科,逐渐构建起了致力于交叉研究的交叉学科。计算社会学是社会学与计算机科学有机交叉的典型,是在互联网平台、人工智能等技术支撑下构建的一种新型社会学研究方法体系^[19]。计算社会学不仅与传统的系统科学、控制论和复杂性科学交叉,并且跨越社会科学等多个领域,如经济学、生态学、社会网络组织、人口学中的小群体动力学、环境以及城市规划等^[20]。随着大数据的引入,传统社会学研究中面临的数据供给匮乏的客观限制会被克服,同时研究思维也从以计算为驱动转向以数据为驱动,为社会数据获取和分析,互联网模拟实验以及“基于行动者的模拟方法”(Agent-Based Modeling)提供了重要的支持,有效地推动了社会学朝着可计算、可分析的数字化、数据化研究发展路径。

六、并非多余的结论

对于大数据的价值功用,学术界仍有不同的理解,但不论我们怎样理解大数据,大数据时代的到来已成不争的事实,对于社会科学来说,大数据的作用不仅局限于研究方法的革新和重构,更重要的是社会科学研究范式的创新,建立起以数据为核心驱动、以“计算”为关键技术的研究范式^[21]。以大数据为驱动的社会科学研究是现代科学技术与学术研究有机结合的必然趋势,以互联网、移动终端、智能设备为平台支撑,共同为社会科学研究持续供给海量、即时、动态的数据,很大程度上弥补了传统社会科学研究面临的数据匮乏问题、数据时滞性问题以及数据典型性和代表性问题。以各种数据分析及挖掘技术为核心的数据处理技术擅长挖掘数据的关联性,通过与传统的因果分析相结合将更为清晰地解释研究结果,与此同时,大数据时代的一系列分析工具也将帮助社会科学家更为精准和细致地理解和剖析数据的内在涵义以及以何种方式才能更为有效、可信地处理数据^[22]。

尽管大数据所带来的技术红利能够很大程度上促进社会科学研究的跨越式发展,推动社会科学的历史性转变,但是从历史经验来看,新的研究范式代替传统研究范式的过程总是要历经长期的争论和辩驳的艰辛过程,新的研究范式在发展初期必然会遭遇技术、惯性和惰性的阻碍和挑战。不仅如此,由于研究者很难确保数据源是否征得受访者的同意,侵犯公众隐私的潜在隐患也成为一大困扰。再者,大数据技术擅长的是数据收集、分析,偏重于寻求解决社会问题的策略和方法,对于发现社会问题则是它的弱项。

从一定程度上来说,大数据并非万能,有其特定的功用范围和应用边界,大数据在社会科学价值功用的发挥不仅需要来自自然科学领域擅长数理推理和数据分析的学者,也需要社会科学研究者发挥其发现问题和理论框架构建的特长,不同学科间的合作程度互补效果也直接决定着大数据与社会科学融合的契合程度,也影响着大数据驱动下社会科学研究能否完全体现其应有的价值功用,总而

言之,只有通过跨学科、跨领域、跨组织的合作才能真正建立以大数据为核心驱动的社会研究新范式^[23]。

注释:

- [1][美]Jay Lee:《工业大数据——工业4.0时代的工业转型与价值创造》,邱伯华译,北京:机械工业出版社,2015年,第46—47页。
- [2]殷杰、樊小军:《社会科学范式及其哲学基础》,《山西大学学报(哲学社会科学版)》2010年第1期,第1—8页。
- [3][13]Ray M.Chang,Robert J.Kauffman,YoungOk Kwon,“Understanding the paradigm shift to computational social science in the presence of big data”,*Decision Support Systems*,2014(7),pp.67—80。
- [4]Zhiwen Yu,Xingshe Zhou,Daqing Zhang,“Understanding social relationship evolution by using real—world sensing data”,*World Wide Web*,2014(16),pp.749—762。
- [5]Duc T.Nguyena,Jai E.Jung,“Real—time event detection for online behavioral analysis of big social dat”,*Future Generation Computer Systems*,2016(66),pp.137—145。
- [6]Stuart E.Middleton, Lee Middleton,“Stefano Modafferi.Real—Time Crisis Mapping of Natural Disasters Using Social Media”,*SOCIAL INTELLIGENCE AND TECHNOLOGY*,2014(2),pp.9—17。
- [7]沈浩、黄晓兰:《大数据助力社会科学研究:挑战与创新》,《现代传播》2013年第8期,第13—18页。
- [8]Banu onanç Uyan Dur,“Analysis of data visualizations in daily newspapers in terms of graphic design”,*Social and Behavioral Sciences*,2014(51),pp.278—283。
- [9]Nathalie Henry,Jean—Daniel Fekete,Michael J.McGuffin,NodeTrix,“A Hybrid Visualization of Social Networks”,*Transactions on Visualization and Computer Graphics*,2007(6),pp.1302—1309。
- [10]A.Dimoka,P.A.Pavlou,F.D.Davis,NeuroIS,“The potential of cognitive neuroscience for Information Systems Research”,*Information Systems Research*,2011(4),pp.687—702。
- [11]卢晖临、李雪:《如何走出个案——从个案研究到扩展个案研究》,《中国社会科学》2012年第1期,第118—130,207—208页。
- [12]R.Bapna,P.Goes,A.Gupta,Y.Jin,“User heterogeneity and its impact on electronic auction market design:an empirical exploration”,*MIS Quarterly*,2004(1),pp.21—43。
- [14]Robert M.Bond,“A 61—million—person experiment in social influence and political mobilization”,*Nature*,2012(489),pp.295—298。
- [15]R.Bapna,P.Goes,A.Gupta,“Replicating online Yankee auctions to analyze auctioneers’ and bidders’ strategies”,*Information Systems Research*,2003(3),pp.244—268。
- [16]Roman Tilly,Kai Fischbach,Detlef Schoder,“Mineable or messy? Assessing the quality of macro—level tourism information derived from social media”,*Electronic Markets*,2015(3),pp.227—241。
- [17]Sandra González-Bailón,“Social Science in the Era of Big Data”,*Policy & Internet*,2013(2),pp.147—160。
- [18]Josh Cowls,Ralph Schroeder,“Causation,Correlation and Big Data in Social Science Research”,*Policy & Internet*,2014(4),pp.447—472。
- [19][21]罗伟、罗教讲:《新计算社会学:大数据时代的社会学研究》,《社会学研究》2015年第3期,第222—246页。
- [20]孟小峰、李勇、祝建华:《社会计算:大数据时代的机遇与挑战》,《计算机研究与发展》2013年第12期,第2483—2491页。
- [22][23]Patricia White,R.Saylor Breckenridge,“Trade—Offs,Limitations and Promises of Big Data in Social Science Research”,*Review of Policy Research*,2014(4),pp.331—338。

[责任编辑:李本红]