

新计算社会学： 大数据时代的社会学研究^{*}

罗 玮 罗教讲

提要：本文在全面解读、分析新计算社会学现有研究成果的基础上，追溯了新计算社会学产生的历程，将新计算社会学的主题内容划分为五个方面：大数据的获取与分析、质性研究与定量研究的融合、互联网社会实验研究、计算机社会模拟研究和新型社会计算工具的研制与开发。本文以现有研究成果为实例，分别对这五大研究内容或主题进行介绍和分析。新计算社会学的产生会引发一场社会学的“革命”吗？笔者借助库恩的科学革命理论，对此进行了分析和探索。

关键词：新计算社会学 大数据方法 计算机社会模拟 互联网社会实验

新计算社会学(new computational sociology)这一名词在2014年8月美国社会学界举办的“新计算社会学研讨会”上首次提出。在这次会议上，来自美国顶级名校不同学科的学者展示了新计算社会学领域的最新研究成果。在本文中，我们试图回答三个问题：什么是新计算社会学、其产生经历了怎样的历程？新计算社会学包含哪些重要内容？新计算社会学将对社会学研究产生怎样的影响？

* 2014年8月15日，美国社会学学会年会召开前夕，近百名来自哈佛大学等美国顶级名校不同学科的学者聚集在斯坦福大学“计算社会科学中心”(Center for Computational Social Science)，召开了一次名为“新计算社会学”的学术研讨会，会上首次推出了“新计算社会学”这个新的社会学名词。本文作者之一罗玮当时正在美国加州戴维斯大学社会学系学习，参加了此次会议。来自哈佛、MIT、斯坦福等名校的17名学者在会上报告了他们在新计算社会学领域的最新研究成果。这次新计算社会学会议是一次名副其实的跨学科学术研讨会，与会者除了社会学界的学者之外，更多的是来自计算机科学、生命科学、信息科学、经济学、政治学、心理学等众多学科背景的学者。谷歌(Google)和脸书(Facebook)的技术专家也参加了这次会议。笔者在收集相关资料的基础上写成本文，向国内社会学界及时介绍这一最新的社会学发展动态。在本文的写作过程中，匿名评审专家提出了富有价值的批评与修改意见，谨表谢意。文责自负。

一、新计算社会学产生的历程

什么是新计算社会学,目前还没有人给出一个明确的定义。新计算社会学不是一个特定的社会学理论流派,也不是某种特定的研究方法。就现在的情况来看,笔者认为新计算社会学是当代社会学界借助计算机、互联网与人工智能技术等现代科技手段,利用大数据、新方法来获取数据与分析数据,从而研究与解释社会的一种新的范式或思维方式。其目的是要克服既有社会学研究方法的局限与不足,达到对人类行为与社会运行规律的真实认知与科学解释。

笔者认为,新计算社会学的产生是大数据时代社会学发展的必然结果。当代计算机科学、互联网与人工智能技术的发展是新计算社会学产生的基础条件,而社会学家对社会学研究新方法的不懈探索与追求,是新计算社会学产生的内在动力。

社会学从产生、发展到现在,所走过的是一条坎坷不平的道路,正如美国社会学家柯林斯和马科夫斯基所说的,社会学并不是一门不可能的科学,但的确是一门很艰难的科学(柯林斯、马科夫斯基,2006:1-23)。所谓很艰难的科学,主要是指社会学在研究方法上经历了一个艰难探索的过程。

社会学研究方法被作为重要问题进行探索并引起争论开始于20世纪50年代,此时西方社会学的定量研究迅速发展成为主流研究方法。随之而来的是对各种方法的争论。例如,美国纽约1956年召开了一次“社会测量大会”,聚集了一大批当时顶尖的社会科学学者,对社会科学研究方法的发展提出了多种意见与构想。心理学家史蒂文斯(S. S. Stevens)提出社会科学研究需要测量手段的更新(reinvent measurement),社会学家拉扎斯菲尔德(Paul Lazarsfeld)提出需要关注定性研究与定量研究之间的关系问题(Mohr & Ghaziani,2014),等等。这次会议对此后包括社会学在内的社会科学研究方法的发展起到了重要的推动作用,特别是进一步提高了定量方法在社会学研究中的主导地位。

进入20世纪70年代后,由于计算机的发展与广泛使用,以及由此所带来的各种数据分析统计软件的问世,社会学研究在大样本问卷调查、数据的多变量统计建模与分析方面,达到了一个前所未有的水平。

与此同时,人们也在积极探索其他研究方法,如进行社会科学实验和开展社会现象的计算机建模研究等等。

尽管如此,社会学研究方法所面临的问题似乎越来越多、越来越严重。社会学研究方法专家谢宇教授坦承,现在用于研究社会和社会关系的所有方法,包括定量与定性方法,都存在局限性。在他看来,社会学中最明显也是后果最为严重的矛盾是研究方法上的矛盾,社会学也因此而分化为不同的阵营(谢宇,2012:5)。

社会学研究方法所面临的困境,实际上是人类行为研究所受时代条件限制的反映。大数据时代的到来,正在为社会学研究方法突破困境创造条件。

20世纪90年代中后期以来,一系列技术进步使得社会学研究方法的进一步创新成为可能,其中最重要的成就表现在四个方面:其一,社会网络理论与研究方法的发展;其二,人工智能的发展带来新型文本与影音资料处理系统的问世;其三,计算机模拟领域内基于行动者模拟方法(agent-based modeling, ABM)的发展;其四,互联网的快速发展,特别是移动互联网时代的到来。1996年,经济学界先人一步,出版了《计算经济学手册》(*Handbook of Computational Economics*)第一卷(Amman et al., 1996),正式宣告“计算经济学”的诞生。在最近10年内,正是这些新的理念和技术推动着人们不断探寻社会学研究方法上的突破,为新计算社会学的产生做好了必要的准备,新计算社会学的孕育过程逐渐完成。

2009年,包括哈佛大学教授拉泽尔(Lazer et al., 2009)在内的15名顶级学者在《科学》(*Science*)上共同署名发表论文,正式提出“计算社会科学”(computational social science)这一概念。文章预言一个以新电脑技术、互联网为基础,具有无限可能性的计算社会科学的产生正在成为现实,甚至在谷歌、雅虎这些大型企业内,已经开始了计算社会科学的研究。过去,我们只能够获取间断的、片面性的社会数据,而如今,借助视频监控、电子邮件、计算机智能命名系统等,社会科学家搜集与处理海量数据的能力得到了空前提升,这正是计算社会科学得以产生的一个重要原因。另一个原因是认知科学的发展。人类对自身认知机制的深入了解,神经生物学、计算机科学以及其他学科的融合,为人类行为研究的计算机模拟提供了条件。拉泽尔教授等人的论文列举了一系列计算社会科学可以大显身手的研究领域或研究对象,由此展示和

证明计算社会科学的魅力与发展前景。这些典型例子包括:利用视屏监控设备记录婴儿最初两年的所有成长数据,以此来研究婴儿的成长机制;通过收集人们的电子邮件数据研究人们的互动行为;利用“社会测量计”(sociometer,一种形状类似身份卡,可以别在胸口,能够捕捉一定范围内行为者的空间位置、互动情况等信息的电子检测仪器)和手机上的GPS记录软件进行人们行为的时空社会学研究;通过社交网站来收集数据,对每个人的健康、心情与品位及行为方式等众多变量进行测量;通过互联网收集美国政治竞选时期谣言传播的数据,分析谣言传播的模式,特别是可以利用互联网进行以前无法进行或不被允许进行的社会实验;大量的文本资料将能够被数据化……这些新技术的应用,将使得经济学、社会学、政治学等社会科学的研究进入一个新的时代。

在2009年发表的论文中,拉泽尔等人尚认为,所谓的计算社会科学只是由一小部分顶尖的计算机专家、物理学家与社会科学家在共同推动的勇敢事业。但5年之后的今天,拥有这种勇敢精神的社会科学家已越来越多,2014年“新计算社会学研讨会”的召开就是例证。笔者相信,这个由不同学科的社会学研究者共同推动,综合运用当代计算机和互联网及其他高新技术与大数据分析手段的社会学研究方法体系,将会越来越受到社会学学者们的关注,并将有可能成为未来社会学发展的主流。“新计算社会学”中的“新”,其一是指新计算社会学在理念、方法、思路、工具应用等方面比此前的社会学研究中的“计算”都更为先进和复杂。其二是“计算社会学”(computational sociology)这个名词已在瑞泽尔(George Ritzer)2007年出版的《布莱克威尔社会学百科全书》(*Blackwell Encyclopedia of Sociology*)中出现(Bainbridge, 2007),只是该书中“计算社会学”词条的内容与“新计算社会学”不同,为了与之区别,2014年的斯坦福会议才提出“新计算社会学”。

从“计算社会科学”概念提出到“新计算社会学”新名词问世,中间经历了5年时间(2009-2014)。在这5年中,社会学家受到了计算社会科学的启发和影响,但鉴于社会学在社会科学中的独特性,社会学家认为社会学有必要脱离计算社会科学而自立门户。所谓社会学的独特性,是指社会学的研究对象、理论视角和研究方法的要求等与其他社会科学如经济学、政治学存在区别。“计算社会科学”与“新计算社会学”中核心内容都是“计算”,其重点都在于计算机科学、互联网与人工智

能技术等的有效应用,只是二者在面对不同研究对象、不同理论解释框架和不同研究方法要求的条件下,“计算”的理路、方法与工具存在差别。

二、新计算社会学的五大内容

新计算社会学的目标是借助各种与社会学研究相关的新技术、新工具、新手段,克服以往社会学研究中存在的各种缺陷与障碍,提高社会学研究的科学性与有效性,开创社会学发展的新时代。要实现这个目标,必须实现社会学研究各个环节、各个方面的创新,因此新计算社会学实际上是一个全面创新的社会学研究方法体系。根据对现已发表的论文和在会议上展示的研究成果的分析,我们将其划分为五个互相关联的组成部分:大数据的获取与分析、质性研究与定量研究的融合、互联网社会实验研究、计算机社会模拟研究和新型社会计算工具的研制与开发。

(一)大数据的获取与分析

数据、资料的获取与分析,是社会学研究的两大关键问题,也正是在这两个环节上,社会学研究受到的批评和诟病甚多。大数据的获取与分析,有望为解决问题找到新的突破口,例如金(King, 2009)提出,未来的研究可以从文本内容、选举活动、商业行为、地理位置、健康信息等数据着手,通过大规模与时序性数据的研究改变政治学乃至社会科学的基础。

范德里特等(van de Rijt et al., 2013)关于“名气”的研究是在《美国社会学评论》(*American Sociological Review*)上最早发表的一篇社会学大数据研究论文。研究者发现,在名气等级体系中,处于底层名气最小的人们确实非常容易被人们遗忘,但在名气体系的上层,即便是在娱乐、电视、博客等看起来非常“健忘”的领域内,个人名气一旦建立,名气的自我增强、职业地位、纪念活动等都能保护个人名气影响力的长期存在。他的研究方法是从报纸上获取个体姓名出现的次数,出现次数越多,代表该人的名气越大(当然,作者也承认这种测量方法本身具有一定的局限性,一个人的名气也有许多是通过其他形式例如影音资料

或者民间口传的形式表现出来的)。作者借助“莉迪亚文本分析系统”(Lydia text analysis system)作为研究工具。该系统利用其“自然语言处理系统”(nature language processing)将文本变成时序化的数据并进行定量分析。作者分析了2004-2009年的2200种美国各类日报与周刊,既有全球性媒体如《时代》(*Time*)周刊,也有地区报纸如《萨克拉门托报》(*Sacramento Bee*),期刊的内容更是遍及要闻、政治、体育、时尚、娱乐等各个领域,从这些报刊杂志中,作者提取出了10万个姓名。为保证数据的合理性,删除了大量过于常用的姓名(以保证不会有同名人士多次出现引起测量干扰),也消除了虚构人名、去世人的姓名等干扰。在这10万个姓名中,大多数的名字在媒体上只出现过几次,但也有上百人的姓名出现了1万次以上,从而形成了一份包含姓名及其每天出现次数的“大数据”。接下来,范德里特等对这些人名进行了分布分析与流动性分析,他的数据处理过程并不复杂,只用到了非常简单的社会学统计分析方法,但他的数据搜集与整理的工程却非常繁琐,从几乎全国所有杂志中提取和统计人名,在大数据时代到来之前是无法想象的。

大数据社会学研究所采用的数据量远大于传统的实证社会学研究,与上文相似,加州大学伯克利分校教授梁(Leung, 2014)的求职网站进行了一次研究,分析了某求职网站上2000-2004年16569名自由职业者对119648份职业提出的964034次申请;萨韦德拉等(Saavedra et al., 2013)关于股票交易员行为模式的研究更是实时记录了2007年1月1日-2008年12月31日这两年间30万次详尽的股票交易记录;乌兹等人(Uzzi et al., 2013)与古德和梅西(Golder & Macy, 2011)的研究,其数据量达到数千万乃至数亿;米歇尔等(Michel et al., 2011)利用谷歌图书进行的文化分析,研究了1800-2000年200年间英语世界文化的流变趋势,其数据库甚至覆盖了人类所有出版书籍的4%,是大数据研究的里程碑式的成果。

从以上代表性文章来看,大数据与传统数据的区别主要在于三个方面:第一,传统数据样本量一般较小,而大数据论文则动辄数十万、上百万,大数据环境下,样本几乎等于总体,研究者甚至没有进行抽样的必要。第二,传统数据常用问卷调查方法获取,数据主观性高、可信性低,而大数据论文所采用的基本上是“自然数据”,这些数据并不通过问卷获得,而是在现实生活中自动形成,可信度大于传统问卷调查数据。第三,传统数据的产生过程是“搜集”,设计问卷后进行调查,问卷

的针对性强,但问卷的应用范围受到限制,为一个研究而进行的问卷数据搜集很难很好地应用于另一项研究,而大数据社会学研究则重在数据的“挖掘”,客观数据并不为任何一个课题而产生,而是对真实世界的自然记录,有利于研究者充分发挥社会学的想象力,可以挖掘的数据无穷无尽,可供研究的领域没有边界。

(二)质性研究与定量研究的融合

如何更加有效地利用文本、影音等质性资料开展研究,是社会学期以来面临的难题。有效研究方法的缺乏,造成了质性研究与定量研究之间一直无法弥合的鸿沟。大数据时代的到来,为社会学的发展提供了更加有效的研究方法与研究工具,使定量研究与定性研究的融合成为可能。刘易斯等(Lewis et al., 2013)更提出,大数据环境下,计算机方法与人工处理的混合使用,可以让传统的“内容分析”方法得到升华。

在发表于《科学》(*Science*)上的一篇论文中,古德与梅西(Golder & Macy, 2011)研究了人们每天或者每个季度的心情变化。要研究人们心情的变化趋势,通过对传统数据的统计分析能力有不逮。研究者们通过大数据的应用与文本资料的量化分析很好地解决了这个问题。两位研究者通过数据挖掘技术从社交网站推特^①(Twitter)中获取了来自82个国家240万人多达5.09亿条英语推特文。然而,推特文是文本资料而非数据,计算机可以识别却无法进行分析。研究者使用“自动文本分析系统”(linguistic inquiry and word count, LIWC)对推特文进行内容分析。LIWC系统是最先进的文本内容分析软件之一,通过对文本的分析,它能够识别出包括焦虑、愤怒、抑郁等在内的60多种人类情绪,并进一步将其归纳为“积极情绪(PA)与消极情绪(NA)。通过这样的内容分析,研究者们能够从每一条推特文中识别出该段文字发表时作者的积极或消极情绪。分析结果发现,人们在早上起床时的心情一般都很好,但在一天中会逐渐发生变化;人们在周末更加快乐,但是这种快乐峰值到达的时间要比工作日晚两个小时;人们的情绪会随着季节的变化而发生变化等等。

① 推特是在许多国家非常流行的一种社交工具(类似微博),人们可以在自己的推特账户上实时记录自己的心情并发表(每篇twitter限制在140个字以内),这些“推特文”可以看作是每个人的心情日记,具有很高的真实性。

与此类似,麦凯尔维等人(McKelvey et al., 2014)也运用同样的方法对3万名用户的11万条推特文进行分析,研究了网民关于国会选举的讨论与他们线下政治行为之间的动态关系;乌兹等人(Uzzi et al., 2013)则对发表在各类学术期刊上的1790万篇文献的质性内容进行了定量分析,探索学术创新的规律及其影响因素,得出了颇有价值的研究结论。此类研究的另一个特点是“可视化”方法的成功应用,许多质性数据并不一定要进行复杂的模型研究,往往只需要实现简单的可视化处理,就足以揭示出非常重要的现象与规律。

定性研究与定量研究融合的关键是文本资料分析工具的研制与开发。从现在的情况来看,虽然这个领域的研究工作还刚刚起步,尚有很多问题需要解决,但其发展的速度很快,在短短的几年时间内已经研制、开发出多种文本内容分析软件系统工具。随着该领域研究工作的步步深入,各种更为先进、精细的文本分析工具(包括中文分析工具)会不断问世,真正实现定性研究与定量研究的融合只是时间问题,由此引发的将不仅是研究方法上的创新,更为重要的是导致人文社会科学研究理念和思维方式上的变革。

(三) 社会学互联网实验研究

社会学的研究方法体系中早就有实验方法的位置,而且也有运用实验方法开展社会学研究的先例。但社会学界对实验方法一直存有戒心,因为运用实验方法来研究社会现象的确存在诸多难以克服的弊端和障碍。运用互联网这个平台来进行社会学的实验研究,是一种创新,而且有可能使实验法成为未来社会学研究的主流方法。

萨尔甘尼克等人(Salganik et al., 2006)进行的关于文化产品市场不公平性的社会实验开启了互联网社会实验研究的先河。他提出一个问题:文化产品在市场上的流行程度受什么影响?一种观点认为市场中产品的质量决定它是否受欢迎;另一种观点则认为市场是不公平的,存在着“超级巨星效应”或者“赢者通吃”(winner-take-all)现象,因为人们对文化产品的选择受到其他人选择行为的影响。为了研究这个问题,研究者在互联网上邀请了14341名参与实验者,为他们提供48首从未面世的歌曲,并邀请这些参与者根据自己的喜好为这些歌曲打分或下载。参与者分为两组(实验组和控制组),实验组能够浏览别人对歌曲的评价,控制组则不能了解其他人对歌曲的评价。实验歌曲按随

机排列的顺序同时提供给实验组和控制组。实验结果发现,在控制组中受欢迎的歌曲在实验组中更加受欢迎,表明人们对歌曲的评价受到了他人评价的影响,即存在所谓“赢者通吃”的现象。在更进一步的实验中,实验组成员看到的歌曲不再随机排列,而是根据下载量排列,控制组依然不能看到他人的评价,他们看到的歌曲依然随机排列。实验结果显示,实验组和控制组对受欢迎歌曲评价的差别进一步扩大。这个实验研究得出结论:首先,对文化产品而言,质量是关键,因为无论是控制组(每个人单独对文化产品进行评价)还是实验组(每个人在参考他人评价的情况下选择文化产品),质量好的文化产品都更加受欢迎。其次,“赢者通吃”的现象确实存在,在他人选择行为的影响下,控制组中受好评的产品在实验组中更受好评,这说明文化产品市场上确实存在不平等性。

与此相似,麻省理工学院的森托拉(Centola,2010)对1528名实验参与者进行的社会网络中行为传播的研究、特维克瓦与梅西(Tsvetkova & Macy,2014)进行的关于“慷慨”的社会“传染”实验都是社会学互联网实验研究的上佳之作。范德里特等(van de Rijt et al.,2014)关于个体成就的互联网实验更是通过四个公众网站,进行了四场社会学实验,成功地实现了互联网实验的“组合拳”研究。

社会学互联网实验是一种全新的实验方法,是真正理想的在自然条件下进行的社会实验研究。这种实验研究的优势在于,它不仅消除了传统社会实验研究存在的某些弊端,如“霍桑效应”等,尤其是不受时间和空间的限制,这为社会学提供了通过互联网实验研究人类行为与社会现象的无限可能性,设计巧妙的互联网社会实验将会极大地扩展和丰富社会学家的社会学想象力。

(四) ABM 模拟方法在社会学研究中的运用

社会学的计算机模拟研究方法已经发展到第三代,即“基于行动者的模拟方法”(agent-based modeling,ABM)。

最早的社会学计算机模拟研究产生于20世纪60年代,其理论基础是结构功能主义学说,重视的是对宏观变量如组织、企业、城市、人口发展变化等的模拟,也就是在历史数据的支持下,模拟宏观社会现象的演化路径。从20世纪70年代开始,微观模拟逐步取代宏观模拟。研究者通过对微观个体行为的观察与测量获取数据,由此对个体的行为

进行演化模拟与推测,了解个体行为决策的机制(Macy & Willer, 2002)。20世纪90年代后,第三代社会学计算机模拟——基于行动者的建模方法由阿克塞尔罗德(Axelrod, 1997)所进行的计算机模拟囚徒困境全球竞赛首开先河。他在全世界邀请多学科专家编写以囚徒困境为博弈规则的计算机竞赛程序,让这些计算机程序进行博弈,以博弈的收益高低(得分多少)计算成败。竞赛结果,一个在所有程序中最短小精悍(一共只有4行程序)被称为“一报还一报”(tit for tat)的程序获得冠军。此后问世的“人工股市模拟”(artificial stock modeling)(Arthur et al., 1997)更进一步,不仅利用计算机程序模拟人的行为,更让程序具有了自我学习的能力,使之更加接近复杂与互动过程中不断变化的真实世界,因此开创了基于行动者(agent)模拟方法在行为金融学中应用的新领域。哈平(Halpin, 1999)、吉尔伯特(Gilbert, 1999)、勒巴龙(LeBaron, 2011)等也先后对新型计算机模拟方法在社会科学研究中的运用做出了贡献。此外,梅西与威勒(Macy & Willer, 2002)还对ABM模拟方法进行过深入的理论探索和系统总结。我国学者对ABM在社会科学中的应用也早已开始,如沙莲香教授等人(2007)曾经运用ABM方法进行过社会心理学问题的研究。尤其值得关注的是,计算机模拟可以描述微观层面上的行动者互动如何演化为宏观层面上的社会现象,因而这种方法预示着一种连接微观与宏观、理论与实证的可能性,2005年赫斯特洛姆(Peter Hedstrom)在《解析社会:分析社会学原理》(赫斯特洛姆,2010)中通过一个关于互动与失业的ABM模型,向我们展示了ABM模型的可观前景。

关于ABM方法在社会学研究中的具体运用,曼佐和波尔多萨里(Manzo & Baldassarri, 2014)关于社会地位形成机制的研究为我们提供了很好的案例。在这项研究中,研究者假设,一个人社会地位的获得取决于其他人对他的“礼敬”态度,人们对他越礼敬、越尊重,他的社会地位越高。社会中存在两种互相冲突的机制来决定一个人得到的“礼敬”,分别是制造不平等的“社会影响”(social influence)机制和限制不平等的“相互礼敬”(reciprocation in deferential gestures)机制。“社会影响”机制遵从现实法则,一个人的成就如财富、权力越高,其他人对他的评价就越高,因而对他越礼敬尊重,社会地位也就越高。在这种机制中,对他人的“评价”等于对他人的“礼敬”,因而这种机制追求“公平”而不是“平等”。“相互礼敬”机制则只关注人与人之间的互相尊

重,一个人如果对他人的尊重,他人也会反过来给予他较高的“礼敬”,从而获得较高的社会地位,反之亦然。因此,这种机制中对他人的“评价”与对他人的“礼敬”并不一定相等,人与人的互相尊重能够促进人与人社会地位的平等。

研究者在计算机中构建了包含 30 名行动者的虚拟空间,每名行动者具有不同的内在素质 (intrinsic quality) Q , 代表该名行动者的成就。互动中的行动者对彼此的素质进行评价,得到评价价值 q , 并根据这个评价价值来赋予对方一定的“礼敬” a , 行动者在多次互动中收获到礼敬的总量决定每个人的社会地位 S , 这样的过程重复 N 次“迭代”。

N 次迭代后,如果社会地位 S 与内在素质 Q 高度一致,则说明成就越高社会地位越高,遵从的是公平但不平等的“社会影响”机制,否则说明“相互礼敬”的机制发挥了作用,社会地位分层更趋向于平等。

在第 1 次迭代时,有公式 $q_{ij1} = Q_j + e_{ij1}$, 表示行动者 i 在第 1 次迭代时根据 j 的真实内在素质 Q_j 对 j 给出一个评价 q_{ij1} , 这种评价可能存在误差,因此引入正态分布误差值 e_{ij1} , 用以模拟现实生活中“看走眼”的情况。行动者 i 根据这个评价 q_{ij1} 来等量地对 j 赋予“礼敬”,所有的参与者在本次迭代中对 j 赋予“礼敬”的均值,即为 j 在此轮迭代中的地位。

在接下来的迭代中,30 名行动者将继续进行多次的上述互相评价、给予“礼敬”、确定地位的过程,不同的是,此后的每一次迭代中,行动者都会参考上一次迭代的数据,并以此调整自己的行为。

在第 t 次迭代中, i 对 j 的评价还需要参考上一个迭代中所有行动者对 j 地位的综合评价。 i 对 j 赋予的“礼敬”(用 a_{ijt} 表示)由他对 j 的评价与上一次迭代中 j 对 i 的“礼敬”态度共同决定,如下式所示:

$$a_{ijt} = \begin{cases} q_{ijt} & , \text{若 } ddd_{ij(t-1)} = a_{ij(t-1)} - a_{ji(t-1)} \leq 0 \\ q_{ijt} - s_i \cdot ddd_{ij(t-1)} & , \text{若 } ddd_{ij(t-1)} = a_{ij(t-1)} - a_{ji(t-1)} > 0 \end{cases}$$

上式表示,如果上一次迭代中 j 对 i 有足够的“礼敬”(即 j 对 i 赋予的“礼敬”不小于 i 对 j 的“礼敬”),则“人敬我一尺我敬人一丈”, i 给予 j 的“礼敬”与 i 对 j 的评价相等。如果上一次的迭代中, j 对 i 不够礼敬,则 i 也会“以眼还眼”, i 对 j 的“礼敬”将会小于 i 对 j 的评价。

如此的循环一直到最后一次迭代,此时, j 的地位由每次迭代中所有行动者对 j 的地位的评价按照一定的权重共同决定。

一般情况下,当迭代次数 N 足够大后,模拟的结果将会趋于稳定。这个 ABM 模型模拟的正是人际交往中的人际社会地位形成的机制。一

方面,他人对此人的评价与个人成就直接相关,追求公平而不是平等;另一方面,交往中的“相互礼敬”与“以眼还眼”等规则不关心个体成就而关注个体在交往中的受尊重情况。迭代 T 次之后所有行动者的地位状态若与每个行动者的内在素质高度一致,则说明社会地位是不平等的,若与行动者的内在素质相似性不大,则说明社会地位的分布更倾向于平等。计算机模拟显示,在一个社会中,人与人之间的互动越容易发生,人们越能够遵循“以眼还眼”的原则,则社会地位的分布越趋向于平等。

ABM 计算机模拟方法在研究复杂社会现象的演化过程与变化机制方面,具有其他研究方法所无法比拟的独特优势。随着 ABM 方法的不断完善与成熟,它在社会学研究中的运用会越来越普遍。但它的运用也对研究者的数学能力提出了比较高的要求,有些研究者具有很强的理工科背景,其使用的数学方法更是艰深。例如森托拉(Centola, 2013)关于自增长型群体的研究,甚至使用了物理学中的临界质量公式,读者如果不具备相当的数理能力,几乎无法阅读和理解。

(五)新型社会计算工具的研制与开发

新计算社会学是一个新的社会学研究方法体系,它产生和发展的物质基础是互联网,其支撑条件是计算机、人工智能等新技术。在新计算社会学实现其研究目标的过程中,需要综合运用互联网技术、计算机以及人工智能技术,根据数据获取与分析的要求,开发出能够有效实现研究目标的具体操作工具,我们称之为新型社会计算工具的开发。

新型社会计算工具多种多样,可以根据具体研究的需要进行研制与开发。例如,麦考利与莱斯科韦茨(Mcauley & Leskovec, 2014)开发出一种网络算法,用以检测社交网络用户各类联系人的信息,包括姓名、年龄、职业、学历等。把这些信息与网络用户本人的信息进行对比,通过各种测量相似性的算法,估算联系人与用户的关系,将这些不同的联系人归入不同的组群(如好友、同事、同学等),实现用户个人网络的自动分组。该算法在实验验证阶段已经获得了成功。这个新型社会计算工具的开发更像是为谷歌、脸书等网站进行的改良顾客体验的研究,它如此技术化,以至于许多社会学者几乎无法看懂。但这种研究对社会网络研究与社会网络理论的发展意义重大。运用这种工具,研究者们能够从社交网络媒体中挖掘出大量的社会网络数据,并有可能利用这些数据进行更大规模、更精准的社会网实证分析。类似地,韦斯特等

人(West et al., 2014)开发出了一套文本分析方法,能够对维基百科语料库(Wikipedia Requests for Adminship Corpus)与肯沃特国会演讲语料库(“Convote” Corpus of Congressional Speeches)进行处理,获取大量的社会网络数据。马瑟等人(Mathur et al., 2012)开发出的视频文件内容分析方法能够通过视频与照片文件进行“内容分析”,分析人与人之间关系的亲疏程度,实现人际网络结构的自动识别。

新型社会计算工具的研制与开发,方式与途径多种多样。为了进行某项课题研究,社会学研究者与精通计算机技术的专家合作,可以量体裁衣地开发出研究所需要的某些小型工具。但对于那些大型且功能复杂工具的研制,则需要依赖多学科共同努力,借助专业公司的力量,甚至依靠国家的实力才能完成。

三、总结与思考

新计算社会学是社会学研究发展的最新前沿,当代科学技术的发展为新计算社会学研究提供了各种必备的条件,那就是以计算机技术为基础发展起来的互联网技术和人工智能技术等高新科技手段与研究方法,而其中最主要的是大数据的社会学应用。对于大数据,学术界现在还存在不同的理解,有人认为只要数据的量足够大就是大数据,也有人认为只有收集到了研究总体的数据才算是大数据(样本=总体)(迈尔-舍恩伯格、库克耶,2013:37-43)。无论我们怎样理解大数据,大数据时代的到来是不争的事实。对于社会科学来说,大数据时代所带来的不只是研究方法上的创新,更重要的是新的社会范式和社会科学范式的出现可能引发的一场社会科学革命,有人创造性地把这两种新的范式分别称之为“社会的信息处理范式”(information processing paradigm of society)和“社会科学计算范式”(computational paradigm of social science)(Cioffi-recilla, 2014)。“社会信息处理范式”意指社会已经进入大数据时代,数据信息的收集与分析已经被人们公认为社会中最为重要的事项或行为方式。^① 信

^① 范式(paradigm)是美国科学哲学家托马斯·库恩提出的概念,指的是在一个特定的时刻或时期内规范科学工作的统摄性理论框架。库恩所说的范式专指科学研究领域,“社会的信息处理范式”显然是创造性地运用了库恩的范式概念。

息处理的核心是借助计算机进行的“计算”。因此,社会信息处理范式直接导致了社会科学计算范式的产生。新的社会科学范式,推动社会科学研究由小数据时代进入大数据时代,社会科学由“计量”进入“计算”时代,计量社会学(quantitative sociology)、计量经济学(econometrics)上升为“计算社会学”(computational sociology)与“计算经济学”(computational economics)。

“计算”是“大数据”产生的基础和条件,由于有了计算机巨大的“计算”能力,我们才能够通过各种渠道(如互联网)和各种技术(如人工智能)收集和处理“大数据”。但新计算社会学研究不等于大数据分析。根据美国社会学界已经发表的研究成果,本文把新计算社会学划分为相互关联的五大内容:大数据的获取与分析、质性研究与定量研究的融合、互联网社会实验研究、计算机社会模拟研究和新型社会计算工具的研制与开发。笔者的分类不一定准确,而且随着新计算社会学的成熟与发展,其内容会更加丰富和完善。但从目前的情况来看,这五大内容充分反映出了新计算社会学的特性和发展前景。新计算社会学的核心同样是“计算”,完成“计算”的工具可以分为“基础工具”和“操作工具”两大部分。基础工具包括计算机、互联网和人工智能技术等;而操作工具则指以“基础工具”为工具,根据研究需要开发、研制,用于获取、挖掘数据和分析数据的各种软件和硬件设备与技术,即“新型社会计算工具”。研制、开发新型社会计算工具是开展新计算社会学研究的必备条件,“抓取”数据和分析数据、进行互联网社会实验研究、开展计算机社会模拟,必须开发、研制专门的工具,质性研究与定量研究的融合,更是需要研制出复杂的文本分析系统等工具。

我们从新计算社会学与传统社会学研究比较的角度,能够更好地认识新计算社会学的特性与发展前景。

其一,新计算社会学是从传统社会学的基础上发展起来的,其发展的外在条件是大数据时代的到来,而内在动力则来自于社会学家对社会学科学理想境界的追求。新计算社会学获取与分析数据的四大方面,即大数据的获取与分析、质性研究与定量研究的融合、互联网社会实验研究和计算机社会模拟研究,都是在传统社会学研究的基础上,力图克服其缺陷和障碍,对原有理念与研究方法的拓展与创新。

其二,新计算社会学与传统社会学相比,其跨学科的特性更为突出,特别是社会学和自然科学与技术科学的关系变得尤为重要。这主

要表现为新计算社会学的研究必须开发出“操作工具”即“新型社会计算工具”。新型社会计算工具的开发,需要社会学家、计算机、互联网、人工智能等各方面专家的合作才能完成。这一方面对社会学研究和研究者都提出了更高的要求,但另一方面也使得社会学的科学特色更为突出。

新计算社会学还刚刚诞生,代表性的研究成果也还很有限,全世界都还处于探索阶段。在国外,新计算社会学的概念刚刚提出,但相关的研究早已开始。在国内,陈云松(2015)、孟小峰等(2013)已经对社会学大数据的研究进行了有价值的探索;梁玉成(2015)则提出不应过度迷信大数据;刘军、郭莉娜(2013)利用QQ群进行过网络交换理论的互联网实验;沙莲香等(2007)也通过计算机模拟进行过社会心理学的研究。虽然对新计算社会学的发展前景学术界尚有不同的看法,但它问世的必然性引发了笔者的思考:新计算社会学给社会学带来一场“革命”吗?^① 讨论社会科学革命,我们要借助科学哲学家托马斯·库恩的理论。虽然库恩研究的是自然科学的革命,但他的概念和理论在社会科学研究中也经常被引用。范式就是库恩被引用最多的概念(黄光国,2006:127-135)。虽然库恩没有讨论过社会科学革命的问题,后来的科学哲学家也没有关于社会科学革命的论述,因为社会科学自产生至今没有出现过库恩所说的革命。但笔者认为今天大数据时代不同于以往任何一个时代,新计算社会科学也与以往任何时代的社会科学不同,因此有必要讨论大数据时代的社会科学革命问题。^②

库恩认为任何科学的发展都要经过一个相同的历程:前科学-常规科学-危机-革命-新的常规科学-新的危机-新的革命……这个过程循环往复不断进行,推动科学不断进步。在“常规科学”(normal science)阶段,为“科学家共同体”(scientific community)所公认并用以指导其研究工作的统摄性理论框架或科学研究视角即“范式”形成,科学共同体中的科学家运用相同的范式去解决该领域中的科学难题,库恩称之为“解谜”(puzzle solving)。当常规科学发展到一定时候,科学

① 限于篇幅,笔者在这里只是提出这个问题,并做出简单的分析和思考。笔者认为这是一个值得深入探讨的问题,计划专门撰文进行深入研究。

② 本文讨论的是新计算社会学是否会给社会学带来革命,社会学虽然是社会科学的一种,但由于其特殊性,与其他社会科学如经济学等存在差别,不能够以社会学代表社会科学来讨论革命问题,只能就社会学这门特殊的社会科学来讨论这个问题。

家会遇到既有范式无法解决的“反常”(anomaly)问题,使得人们开始对范式失去信心,于是“危机”出现。在这种情况下,为了解决反常问题,有人会提出新的范式。如果新的范式被科学家共同体广泛接受,科学革命便发生。因此,所谓科学革命,就是一种新的科学范式取代原有范式的过程。一般而言,一种新的范式不可能立即取代原有的范式,总要经过或长或短时间的争论乃至斗争,甚至存在新的范式不能够取代原有范式的情况,那就是科学革命的流产(库恩,2003:19-35)。

作为社会科学的社会学,尽管与自然科学存在区别,但我们仍然认为库恩的理论对解释大数据时代社会学的发展有比较好的解释力。

社会科学与自然科学的最大区别,一是其研究的对象——人,生活在不同国家的政治制度、经济体系、社会文化与意识形态环境之中,因而千差万别;二是社会科学的研究本身也要受到这些因素的影响,因此很难形成全球统一的社会科学。尽管如此,全球社会科学的共性远大于其个性。这是我们讨论社会科学革命的基础。

就社会学而言,“危机”早就出现,既有“范式”不能解释的“反常”社会现象不断出现。社会学家也一直在化解“危机”,提出新的“范式”,但都没有产生“科学革命”,而是形成了多个范式共存的局面,使得今天的社会学成为了一门多范式(multi-paradigm)社会科学(莫家豪,2000:2-26)。^①问题是社会学的多范式共存并没有真正化解“危机”。美国社会学家科林斯和马科夫斯基在总结当代社会学时悲观地感叹:我们所了解的和已经了解的社会世界大部分都是幻象(科林斯、马科夫斯基,2006:1-23)。社会太复杂,要化解当代社会学的“危机”,显然需要一场社会学革命。

大数据时代产生的新计算社会学会引发社会学革命,即大数据时代“社会科学的计算范式”能广为社会学共同体接受和认同,从而取代既有的其他范式或至少成为一种在社会学研究中占主导地位的主流范式,笔者认为这已经是一种必然的趋势。由于社会学学科的特殊性,全球社会学由此形成新的主导与追随关系体系,并以一种全新的方式争夺学术话语权。

^① 库恩的早期理论中认为在科学革命中新范式与既有范式具有“不可通约性”(incommensurability),即新老范式是不相容的。由于受到太多批评,后来他对自己的观点进行修正,认为范式之间可以“部分互通”(partial communication)。可见在自然科学中,也存在多范式情况。经济学、政治学等都是多范式社会科学。

从社会学发展的历史过程来看,新计算社会学所处的大数据时代与以往任何时代都完全不一样。大数据时代与以往的前工业化时代、工业化时代、后工业化时代的最大区别在于,人类社会(从个体、群体、各种社会组织、社区,到国家机器)被计算机和互联网别无选择地纳入一个具有无限空间和行为可能性的网络体系之中,任何人离开这个体系都将无法生存。处于这个体系中的人与组织最重要的生存和发展技能就是信息处理能力。对此,人类社会已经形成了高度一致的共识,这就是“社会的信息处理范式”。在已经形成“信息处理社会范式”的大数据时代,社会学研究最有效、最符合时代潮流的范式只能是“社会科学的计算范式”。如果社会学家此时选择其他范式来开展研究工作,那就意味着“科学范式”与“社会范式”相背离。这样做不是不可以,但其结果可能会事倍功半甚至事与愿违。当然,可以肯定的是,在相当长的时间内,社会学将仍然是“计算范式”与其他范式共存的局面。

大数据时代新计算社会学成为主流后,社会学国际学术话语权的争夺将以一种新的方式展开。

自二战以后至今,以美国为中心的西方社会学一直掌握着世界学术发展的话语权,所依赖的是以美国国力为支撑的社会学研究能力和美国意识形态主导下的社会学计量范式或定量范式(quantitative paradigm of sociology)的全球推广。大数据时代到来之时,社会学的计算范式取代定量范式所引发的社会学革命仍然首先发生在美国,这为美国在大数据时代继续主导全球社会学话语权奠定了非常有利的基礎。但是,新计算社会学时代学术话语权的争夺方式与传统社会学时代有了很大区别。虽然新社会学时代学术话语权的争夺仍然要通过社会学研究的学术产出能力和研究范式的推广能力表现出来,但这两种能力,在新计算社会学时代,都要依赖数据的“计算”能力来实现。数据计算能力的大小取决于“基础工具”与“操作工具”即“新型社会计算工具”的水平与研发能力。因此,未来可能形成的局面是,数据计算水平高的国家可以通过数据垄断——也就是“计算”能力的垄断——来获得学术话语权。新计算社会学对我们中国社会学意味着什么?中国社会学一直在追赶美国社会学或西方社会学,但我们追赶的目的是要建立自己具有本土特色的中国社会学,从而使具有在全世界影响乃至主导学术发展方向的能力。面对新计算社会学时代的到来,我们该如何应对,这是一个值得深入探讨的重大问题。

首先,我们需要突破学术思维“路径依赖”的惯性。客观地说,我们已经形成追随甚至崇拜西方学术发展的习惯性心理定势。虽然新计算社会学在美国也刚刚出现,也还远未能成为主流,但我们必须认识到,随着大数据时代的到来,社会学乃至社会科学“计量范式”向“计算范式”的转换只是一个时间问题。其次,大数据时代的到来为我们中国社会学实现“弯道超车”提供了难得的机遇。我们具有自己的优势,重要的是我们需要认识到这些优势并且充分发挥自己的优势,真正建立起中国社会学人的学术自信。最后,我们必须马上行动起来,整合资源,进行名副其实的跨学科合作研究。

参考文献:

- 陈云松,2015,《大数据中的百年社会学——基于百万书籍的文化影响力研究》,《社会学研究》第1期。
- 赫斯特洛姆,彼得,2010,《解析社会:分析社会学原理》,陈云松、范晓光、朱彦等译,南京:南京大学出版社。
- 黄光国,2006,《社会科学的理路》,北京:中国人民大学出版社。
- 柯林斯,兰德尔、迈克尔·马科夫斯基,2006,《发现社会之旅——西方社会学思想述评》,李霞译,北京:中华书局。
- 库恩,托马斯,2003,《科学革命的结构》,金吾伦、胡新和译,北京:北京大学出版社。
- 梁玉成,2015,《大数据不能替代理性思考》,《社会科学报》2月12日。
- 刘军、郭莉娜,2013,《关系对联盟的影响——基于网络交换论的实验研究》,《社会学研究》第6期。
- 迈尔-舍恩伯格,维克托、肯尼思·库克耶,2013,《大数据时代:生活、工作与思维的大变革》,盛杨燕、周涛译,杭州:浙江人民出版社。
- 孟小峰、李勇、祝建华,2013,《社会计算:大数据时代的机遇与挑战》,《计算机研究与发展》第12期。
- 莫家豪,2000,《社会学与社会分析》,北京:中国社会科学出版社。
- 沙莲香、刘颖、王卫东、陈禹,2007,《社会心理现象计算机模拟及其方法论意义》,《社会学研究》第6期。
- 谢宇,2012,《社会学方法与定量研究》(第二版),北京:社会科学文献出版社。
- Amman, Hans M., David A. Kendrick & John Rust 1996, *Handbook of Computational Economics*. Amsterdam: North Holland.
- Arthur, W. Brian, John H. Holland, Blake LeBaron, Richard Palmer & Paul Tayler 1997, “Asset Pricing Under Endogenous Expectations in An Artificial Stock Market.” In W. Brian Arthur, Steven N. Durlauf & David Lane (eds.), *The Economy as An Evolving Complex System II*. Boston: Addison-Wesley.
- Axelrod, Robert 1997, *The Complexity of Cooperation*. Princeton: Princeton University Press.

- Bainbridge, William S. 2007, "Computational Sociology." In George Ritzer (ed.), *Blackwell Encyclopedia of Sociology*. Blackwell Reference Online.
- Centola, Damon 2010, "The Spread of Behavior in an Online Social Network Experiment." *Science* 329.
- 2013, "A Simple Model of Stability in Critical Mass Dynamics." *Journal of Statistical Physics* 151.
- Cioffi-recilla, Claudio 2014, *Introduction to Computational Social Science*. London: Springer.
- Dohan, Daniel & Martín Sánchez-Jankowski 1998, "Using Computers to Analyze Ethnographic Field Data: Theoretical and Practical Considerations." *Annual Review of Sociology* 24.
- Gilbert, Nigel 1999, "Simulation: A New Way of Doing Social Science." *American Behavioral Scientist* 42.
- Golder, Scott A. & Michael W. Macy 2011, "Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures." *Science* 333.
- Halpin, Brendan 1999, "Simulation in Sociology." *American Behavioral Scientist* 42.
- King, Gary 2009, "The Changing Evidence Base of Social Science Research." In G. King, K. Lehman Schlozman & N. Nie (eds.), *The Future of Political Science: 100 Perspectives*. New York: Routledge.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, Marshall Van Alstyne 2009, "Life in The Network: The Coming Age of Computational Social Science." *Science* 323.
- LeBaron, Blake 2011, "Active and Passive Learning in Agent-based Financial Markets." *Eastern Economic Journal* 37.
- Leung, Ming D. 2014, "Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market." *American Sociological Review* 79.
- Lewis, Seth C., Rodrigo Zamith & Alfred Hermida 2013 "Content Analysis in An Era of Big Data: A Hybrid Approach to Computational and Manual Methods." *Journal of Broadcasting and Electronic Media* 57 (1).
- Macy, Michael W. & Robert Willer 2002, "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28.
- Manzo, Gianluca & Delia Baldassarri 2014, "Heuristics, Interactions, and Status Hierarchies: An Agent-based Model of Deference Exchange." *Sociological Methods and Research* 26.
- Mathur, Shobhit, Marshall Scott Poole, Feniosky Peña-Mora, Mark Hasegawa-Johnson & Noshir Contractor 2012, "Detecting Interaction Links in A Collaborating Group Using Manually Annotated Data." *Social Networks* 34.
- Mcauley, Julian & Jure Leskovec 2014, "Discovering Social Circles in Ego Networks." *ACM Transactions on Knowledge Discovery from Data* 8.
- McKelvey, Karissa, Joseph DiGrazia & Fabio Rojas 2014, "Twitter Publics: How Online Political Communities Signaled Electoral Outcomes in The 2010 US House Election." *Information*,

Communication and Society 17.

- Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, Erez Lieberman Aiden 2011, "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331.
- Mohr, John W. & Amin Ghaziani 2014, "Problems and Prospects of Measurement in The Study of Culture." *Theory and Society* 43.
- Saavedra, Serguei, R. Dean Malmgren, Nicholas Switanek & Brian Uzzi 2013, "Foraging Under Conditions of Short-term Exploitative Competition: The Case of Stock Traders." *Proceedings: Biological Sciences* 280.
- Salganik, Matthew J., Peter Sheridan Dodds & Duncan J. Watts 2006, "Experimental Study of Inequality and Unpredictability in An Artificial Cultural Market." *Science* 311.
- Tsvetkova, Milena & Michael W. Macy 2014, "The Social Contagion of Generosity." *PLoS ONE* 9.
- Uzzi, Brian, Satyam Mukherjee, Michael Stringer & Ben Jones 2013, "Atypical Combinations and Scientific Impact." *Science* 342.
- van de Rijt, Arnout, Eran Shor, Charles Ward & Steven Skiena 2013, "Only 15 Minutes? The Social Stratification of Fame in Printed Media." *American Sociological Review* 78.
- van de Rijt, Arnout, Soong Moon Kang, Michael Restivo & Akshay Patil 2014, "Field Experiments of Success-Breeds-Success Dynamics." *Proceedings of the National Academy of Sciences* 111.
- West, Robert, Hristo S. Paskov, Jure Leskovec & Christopher Potts 2014, "Exploiting Social Network Structure for Person-to-Person Sentiment Analysis." *Transactions of the Association for Computational Linguistics* 2.

作者单位:武汉大学社会学系
责任编辑:杨典