

拥抱数字时代的计算社会学^{*}

——评马修·萨尔加尼克的《计算社会学》

宋庆宇 乔天宇 张樹沁^{**}

摘要: 数字技术使得人们收集与分析数据的能力大大提高,也给社会科学带来了新的机遇和挑战。萨尔加尼克的《计算社会学》一书总结了数字时代社会科学的发展,认为数字技术拓展了社会科学研究想象力与能力,是对传统研究方法的补充,呼吁将传统研究方法与新的研究方法结合起来。不过,萨尔加尼克关于计算社会科学研究方法的讨论忽略了计算模拟的方法。总的来说,萨尔加尼克这本书对于当前处于起步阶段的计算社会科学具有非常重要的指导意义。

关键词: 计算社会学 大数据 计算模拟 研究方法

痕迹数据(digital foot)有相当长的历史,正如舍恩伯格与库克耶(2013:51-66)所说:“计量和记录一起促成了数据的诞生,它们是数据化最早的根基。”但是,由于记录、存储与分析数据能力的限制,社会科学家一直未对痕迹数据进行深入细致的分析。随着数字技术的飞跃式发展,互联网、智能手机、社交媒体、可穿戴设备等信息技术工具出现,实时地监测每一个社会行为成为现实,越来越丰富的痕迹数据产生了。数字时代不仅为社会科学提供了远观社会的望远镜,也提供了更多的计算分析工具与方法,使得社会科学家的数据分析能力大大提高。

数字时代既给社会科学提供了机遇,例如利用海量的新类型数据研究之前无法回答的问题,也给社会科学提出了挑战,例如:如何处理传统调查研究与大数据研究之间的关系?新的数据类型与研究范式对社会科学研究者提出了哪些新的要求?新的研究设计与数据收集在研究伦理上面临哪些新的挑战?马修·萨尔加尼克(Matthew Salganik)的《计算社会学:数据时代的社会研究》(Bit by

^{*} 本文系国家社会科学基金青年项目“数字设备应用中的健康观念塑造机制研究”(项目批准号:20CSH057)的阶段性研究成果。

^{**} 宋庆宇,河海大学公共管理学院;乔天宇,北京大学社会学系;张樹沁,中央财经大学社会与心理学院。

Bit: Social Research in the Digital Age) (简称《计算社会学》)总结了数字时代的社會科学研究,对上述问题进行了回应。萨尔加尼克认为,要将传统研究与新方法研究结合起来,而不是用后者取代前者。

一、萨尔加尼克的“音乐实验室”

马修·萨尔加尼克目前是美国普林斯顿大学社会学教授,研究兴趣包括社会网络和计算社会学,在《科学》(*Science*)、《美国科学院院刊》(*PNAS*)、《社会学方法论》(*Sociological Methodology*)等期刊发表多篇论文。这部《计算社会学》是迄今为止社会学者针对计算社会学做的为数不多的系统性概述。

小世界网络模型的发现者邓肯·瓦茨(Duncan Watts, 1998)是萨尔加尼克的博士指导老师。受瓦茨影响,萨尔加尼克很早就关注数字技术和计算方法对社会科学的影响。在博士求学阶段,萨尔加尼克与彼得·多兹(Peter Dodds)、瓦茨等合作开展了一项在线实验研究——“音乐实验室”(MusicLab),成为应用数字技术进行计算社会学研究的经典。萨尔加尼克通过“音乐实验室”的研究,意识到社会科学在数字时代建立新研究范式的可能性。

随着研究的深入,萨尔加尼克(2019: VII)认识到:“技术的变化,尤其是技术从模拟时代到数字时代的转变,意味着我们可以用新的方式搜集和分析社会数据。”他一直在思索如何利用新的方式开展社会研究,《计算社会学》正是对这一思考结果的总结。在该书中,萨尔加尼克详细介绍了计算社会学的重要研究成果,既包括他自己的相关研究成果,也包括其他学者的研究成果。萨尔加尼克的主要目的是介绍一种新的社会研究范式,因此在书中并没有纠结于具体的技术细节。即便如此,这对于当前处于起步阶段的计算社会科学也是非常及时的。

二、数字时代社会学研究面临的机遇与挑战

从模拟时代到大数据时代,越来越多的数据积累了起来。萨尔加尼克总结了大数据资源的 10 个特征——海量性(big)、持续性(always-on)、不反应性(nonreactive)、不完整性(incomplete)、难以获取(inaccessible)、不具代表性

(nonrepresentative)、飘移(drifting)、算法干扰(algorithmically confounded)、脏数据(dirty)、敏感性(sensitive)。大数据资源并不完美,萨尔加尼克根据大数据与社会研究的关系,将上述特征分为两类,其中前3个特征有利于社会研究,后7个特征不利于社会研究。

萨尔加尼克指出,正是由于大数据资源不完美,因此为了推进社会科学研究的发展,才需要了解新数据资源的特征与优缺点,掌握如何从新的数据资源中获得有用信息。对于社会科学来说,大数据改变我们理解与组建社会的方法,带来思维方式的转变,也给社会科学研究带来了挑战:不仅要面对新的数据源,还要拓展与转换社会科学研究中的想象力与技能(邱泽奇,2018;舍恩伯格、库克耶,2013:1-29、51-66),与对社会行为感兴趣的数据科学家开展合作,以应对当今社会面临的一系列问题。

(一) 机遇

萨尔加尼克强调计算社会学蕴含潜力的同时,并没有一味鼓吹大数据比传统定量数据要好,而是认为不管大数据具有什么特征,其本质上还是数据,是人们社会生活的痕迹(邱泽奇,2018)。作为一种新的数据类型,大数据对调查数据的挑战取决于它对调查数据的替代程度与扩展程度。正如该书中文推荐语所说,“站在数据科学与社会科学十字路口探索计算社会科学的‘美丽新世界’”,数字技术为社会科学研究提供了许多机遇。

第一,大数据意味着“我们已经从一个缺乏行为数据的世界进入一个行为数据极其丰富的世界”(萨尔加尼克,2019:5-8),有许多数据资源“现成品”。首先,互联网企业产生并积累了大量的行为数据,例如搜索引擎中的日志和社交媒体上的帖子;其次,随着二维码、可穿戴设备、感应器等技术设备的普及,企业与政府组织积累了越来越多的线下数据,记录人们日常生活的方方面面;最后,政府的行政记录与统计记录等数据的数字化也成为一个重要的大数据来源:丰富的数据资源为研究者解决很多理论问题提供了可能(Farber, 2015; Ansolabehere & Hersh, 2012)。

第二,日常生活的数据化——社会与个体的日常活动以数字化的形式呈现并存储下来(萨尔加尼克,2019:5-23)。数字技术与设备已经进入人们日常生活的各个角落,如宋庆宇等(2019)揭示了可穿戴设备实时记录着人们日常活动

的地理位置、心率、步数等,使得现代社会生活的数据化特征愈发明显。技术设备使得“社会之镜”(socioscope)成为现实,可以全方位、高精度地观察社会生活,为科学研究提供了新的领域(彭特南,2015:11-19)。

第三,数字设备的互联——移动终端、交易平台、物流网络聚集在一起,形成一个全面监测的环境。数字世界借由其连通性将分散在不同地区的行动者链入网络中,各种数据的汇流让数据具有了社会意义(邱泽奇,2018;邱泽奇、范志英、张樹沁,2015),也为社会研究提供了新可能。例如,实施随机化实验:一方面,人们许多行为都在互联网上进行,因此可以在互联网上持续开展随机对照实验,突破了传统社会科学中实验方法在时间上的约束(Restivo & Rijt, 2014);另一方面,通过数字传感器测量和记录,研究者可以在线下实施对研究对象的随机化处理,实现理想社会科学中的实验标准(Salganik, Dodds & Watts, 2006; Schultz, Nolan & Cialdini et al., 2007)。另外,连通性也创造了新的交流途径,让研究人员可以与世界上其他研究者进行大规模协作,开展创新性调查研究,解决以往研究无法解决的问题。

(二) 挑战

萨尔加尼克(2019:10-15)不仅总结了新技术与新数据带来的机遇,也列出了其带来的挑战:研究人员要全面认识计算社会学的优势与劣势,权衡数字时代社会研究的机会与风险,才能抓住新的发展机会。总的来说,数字时代社会研究面临以下三个方面的挑战:

第一,数据问题。正如上文所言,虽然数字时代的大数据具有海量性等有利于社会研究的特征,但也有许多不利于社会研究的因素。首先,大数据往往掌握在政府与企业手中,社会科学研究者很难获得。社会科学研究者不再像传统模拟时代那样掌握研究设计的主动权,只有与企业、政府等合作,才能接触与分析大数据,失去了对社会研究的主导权(邱泽奇,2018)。其次,数据不易获得与共享。即使研究者与政府和企业合作,但严格的法律、商业机密和伦理也限制了研究人员获取数据。企业尤其将数据视为重要资产,再加上为了规避信息泄露风险,通常不愿意与研究人员合作。即使能够获得数据,但由于数据无法公开与共享,其他研究者也“无法验证和扩展你的研究成果”(萨尔加尼克,2019:36)。最后,数据的代表性问题。根据传统调查标准,计算社会科学处理的数据是非概率

便利抽样的结果,未经任何调整的数据可能会得出糟糕的结论,尤其是将研究结果推断到目标整体的时候,这个问题表现得尤为明显。研究者在尝试结合传统数据与大数据提高数据代表性的时候,会遇到数据稀疏、大数据的质量很难评估等问题。此外,由于数据收集中存在用户飘移、算法干扰等现象,大数据往往与社会研究需要之间存在一定距离。

第二,理论研究。萨尔加尼克认为,利用大数据进行描述性研究虽然重要,但更强调理论在大数据研究中的重要性。正如瓦茨(Watts, 2014)批评社会学一直以来忽视对社会趋势的预测,认为社会科学作为一门科学,其解释必须符合能够预测未来的科学标准(陈云松、吴晓刚、胡安宇等,2020)。但是,计算社会科学在理论解释与预测方面也面临一些挑战:一方面,理论构念操作化的问题依然存在,更多的数据并不意味着构念效度问题迎刃而解,如何更好地对概念进行操作化依然需要讨论与完善;另一方面,计算数据中存在着述行性问题,即“当在线系统的设计者意识到社会理论的存在并将其录入系统的运行方式中时,就会产生更为复杂的算法干扰”(萨尔加尼克,2019:45),数据的干扰性可能是不可见的。当“大数据资源似乎印证了社会理论的预测时,我们必须确保理论本身并没有被纳入系统的运作方式”(萨尔加尼克,2019:44-46)。

第三,道德伦理。在数字时代的社会研究中,研究者面临着艰难的伦理决策,这也是萨尔加尼克在该书中一再强调的主题。技术进步使得研究人员对个体生活的控制日益增强,“在未征得参与者同意,甚至在他们不知情的情况下,对他们做一些事情”(萨尔加尼克,2019:10-13),研究能力增强的速度超过了法律法规更新的速度。

总之,在数字时代,对于应当如何使用新的研究能力,目前学界还缺乏共识。造成这种混乱的根源有很多。首先,计算社会科学没有统一的研究伦理。计算社会科学的研究者来源不同,因此有不同的研究伦理,围绕隐私权等有很大分歧。其次,社会文化规范与法律制度的差异。同样的研究设计在有些国家或地区符合法律规定,但在另外一些国家或地区则是违法的,这个困境在全球化时代更加突出。最后,数据的二次使用。“一个数据库是出于某一种目的而建造的……但有一天它也可能被用于一种截然不同的目的”,这会带来意想不到的后果,即使对数据采取匿名化等措施,也很难完全消除数据中存在的风险(Selt-

zer & Anderson, 2008: 273–328; Barbaro, Zeller & Hansell, 2006; Zimmer, 2010; Narayanan, Huey & Felten, 2016: 357–385)。

三、应对之策:对机遇与挑战的回应

社会科学如何迎接数据科学与社会科学交叉——这个“美丽新世界”——的机遇？萨尔加尼克(2019:13–15)在《计算社会学》中介绍了新研究范式方向，指出计算社会学为传统社会研究带来的变化，“将整合过去我们所做的以及未来将赋予我们的截然不同的能力”，并且回顾、总结了相关的研究。萨尔加尼克并不主张新的研究范式将取代传统社会科学研究范式，而是认为应当“将现成品与定制物结合起来”(萨尔加尼克,2019:10–13)。

(一) 观察行为

在不干扰研究对象的前提下，观察行为持续运作，收集个体、企业和政府日常活动所产生的副产品——行为数据。对观察得到的数据，计算社会科学可以通过计数、预测和近似实验三个主要研究策略获取有用信息(表1)：

表 1 观察行为的三种研究策略

研究策略	内涵	重要性	实证研究
计数	对研究问题进行简单的计数	帮助决策者决策；研究“社会系统如何运作”	Farber, 2015
预测和临近预测	试图利用预测出的观点了解世界的现状	政府等及时准确地了解情况	Ginsberg Mohebbi & Patel et al. , 2009
近似实验	现实世界中(几乎)刚好能随机研究某些人的事情	可以进行因果推断	Mas & Moretti, 2009

萨尔加尼克认为，数字时代的观察行为对于社会研究的价值表现在三个方面：第一，检验以往无法验证的理论解释，尤其是相互竞争的理论；第二，大数据的预测能力给决策者提供了更好的评估信息；第三，在不开展实验的前提下进行因果推断。萨尔加尼克强调，“通过仔细积累经验事实、实际模式和难解之题，研究人员可以建立新的理论”(萨尔加尼克,2019:72)，扎根理论在数字时代有新的可能性，平衡数据和理论之间的关系。

（二）提问

数字技术不会取代传统社会调查,反而可以将传统调查中的提问方法应用到新的大数据研究中,提升调查数据的价值。萨尔加尼克在回顾抽样调查历史的基础上,发现技术和社会变迁驱动着社会调查研究不断变化。在数字时代,新的社会背景需要社会科学研究者认识到改变调查研究方式的必要性。

首先,利用新技术让传统调查形式多样化。信息技术通过语音会话、短信发送微观调查等方式,可以收集更高质量的数据(Schober, Conrad & Antoun et al., 2015)。其次,利用新技术设计新调查、提问方法。相较于传统调查,新技术可以使得研究问题更开放、更有趣、更贴近生活。例如,生态瞬时评估法^①分解了传统的调查内容,将之融入调查对象的日常生活中,得到具有很高研究价值的高频纵向数据。再次,利用新技术发展与完善非概率抽样调查研究。由于调查技术和社会因素的制约,主流的概率抽样调查的无回答率一直在上升(National Research Council, 2013),维持高回答率的成本越来越高,无回答现象对研究结果的可靠性产生重大影响。事后分层技术^②等调查技术使得非概率抽样方法发展到2.0阶段,用更低成本、更快速度来满足调查研究的需要。最后,利用新技术拓展社会研究领域。丰富型提问^③、扩充型提问^④等方法将调查数据与大数据结合起来,实现单独通过调查或大数据无法实现的研究目标。当然,萨尔加尼克也指出,新的研究范式并不保证显著降低研究误差。即便如此,社会科学研究者也应当拥抱新的发展趋势,从传统社会研究范式中吸取经验,推动调查提问方法的进一步发展与完善。

（三）开展实验

如上文所说,实验是萨尔加尼克开始其计算社会科学研究的起点。萨尔加

① 利用生态瞬时评估法,研究人员可以在合适的时间和地点进行提问,而不是在事情发生数周后才通过一个长时间的访谈来了解(萨尔加尼克,2019)。

② 事后分层的主要思想是利用关于目标总体的辅助信息,帮助改善源于样本的评估结论。研究者在利用事后分层从非概率样本中得出评估结论的过程中,将抽样总体分为不同的小组并评估情况,然后利用各组的评估结果计算出加权平均数,进而得出最终的评估结果(萨尔加尼克,2019)。

③ 在丰富型提问中,大数据资源中含有研究人员感兴趣的核心数据,而调查数据围绕该大数据资源构建起了必要的背景(萨尔加尼克,2019)。

④ 在扩充型提问中,大数据资源中没有研究人员感兴趣的核心数据,但研究人员可以用它来扩充调查数据(萨尔加尼克,2019)。

尼克十分重视在数字时代开展大规模实验的潜力:互联网能够突破组织实施方面的限制,进行大型群体实验。传统社会研究中,由于社会现象概念的操作化与测量的困难、社会现象的成因复杂以及伦理等原因,社会实验难以执行。因此,社会科学家往往通过观察和调查来收集资料,通过定量或定性分析方法来探索因果关系。但是,受制于数据与分析方法,因果推断没有实现充分的发展(彭玉生,2011;陈云松、吴晓刚、胡安宁等,2020)。

数字时代为“因果推论的唯一可靠方法”的发展创造了可能。萨尔加尼克从“实验室 VS 实地”“模拟 VS 数字”两个维度将实验分为四种类型。在纯实验室实验中,数字技术可以精确地测量参与者的行为。例如,已有大量研究者在线招募实验参与者,记录参与者在实验过程中的所有操作行为(Bohannon, 2016)。

数字技术也为实地实验带来了更多可能性,“使研究人员能将实验室实验的严格控制和过程数据与实地实验更加多样化的参与者和更自然的实验环境结合起来”(萨尔加尼克,2019:144),解决传统实验难以解决的问题。对社会科学研究来说,这是最有发展潜力的实验方法。首先,数字实地实验能以零可变成本获得数据,拥有数百万名参与者,大大拓展了实验的规模。其次,数字实地实验不间断地记录整个实验过程,获取参与者的背景信息,进行更加高效的实验设计与分析。最后,数字实地实验突破时间限制,实现更长时间跨度的实验设计。此外,数字时代的实验设计在效度、异质性与原理^①三个方面正在经历从简单实验到丰富实验的突破,为实验方法的发展打开了空间。

萨尔加尼克也介绍了数字时代实验方法的缺点或不足。例如,数字时代的实验方法还是不能进行历史研究,只能评估可操控的处理产生的效果(Banerjee & Duflo, 2009; Deaton, 2010)。

(四) 进行大规模协作

“进行大规模协作”指利用微小的努力来进行大规模的科学实验,解决以往似乎不可能解决的问题。万物互联将不同的社会个体连接起来,使大规模的科

^① 数字时代的实验更易于通过实证来解决外部效度的问题,同时也更易于确保内部效度,而概念效度的问题则可能变得更具挑战性。在大多数模拟实验中,研究人员关注平均处理效应,因为只有少量参与者,并且对其了解很少;而在数字实验中,通常会有更多的参与者,研究人员可以分析异质性效应,对研究对象也有更多了解。数字实验能以两种方式帮助我们找出原理:(1)能够搜集更多的过程数据;(2)能够测试许多相关的处理。

研协作成为可能,更多人参与到研究中,发挥不同人的长处,给社会研究带来帮助(萨尔加尼克,2019:199-248)。

大规模协作在天文学(Marshall, Lintott & Fletcher, 2015)、生态学(Dickinson, Zuckerberg & Bonter, 2010)等领域有着丰富而悠久的历史,只是在社会科学研究中很少有研究者使用这种方法。萨尔加尼克认为,在社会科学研究中,大规模协作也有着广阔的未来。大规模协作不仅可以帮助研究者更好地开展研究,还可以节约研究成本。萨尔加尼克根据大规模协作在社会研究中的应用情况,将其大致分为三类:人本计算、公开募集和分布式数据采集。为了推进这个新方法在社会研究中的应用,萨尔加尼克给出了实际操作的五个一般性规则与两个设计建议。^①

(五) 小结

萨尔加尼克指出,新的技术在研究者与研究对象之间建立新的联系,提供新的研究视角与范式,解决社会科学以往难以回答的一些问题。萨尔加尼克主张将新的数据、新的方法与传统社会研究方法结合起来。这是因为,计算社会学研究方法虽然具有速度和成本上的优势,但也存在数据代表性等方面的局限。将这两种研究方法结合起来,能够很好地弥补彼此的不足,发挥数字技术的研究潜力,更好地解决社会问题。

除了介绍社会科学面对新的机遇以及应对的策略,萨尔加尼克还特别指出,社会科学研究的道德伦理问题在数字时代显得更加突出,道德伦理建设具有紧迫性与必要性。建设的现实途径是从已有的道德原则和框架中吸取智慧,结合数字时代社会研究的新特点,形成新研究范式的道德伦理框架。萨尔加尼克重点强调了,计算社会科学有四项基本原则需要遵守:对人的尊重原则、有利化原则、公正原则和对法律与公共利益尊重的原则。然而,在实际研究实践中,将以上四项原则全部落实并不是一件简单的事情,而需要研究者多方面权衡。此外,

^① 五个一般性原则:激励参与者(金钱、个体价值与集体价值等)、利用异质性(技能、努力程度等)、集中注意力(将参与者注意力集中到最有价值的地方)、允许惊喜(允许参与者交流,创造出预想之外的结果)、合乎伦理道德。两个设计建议:“应该先从所有能连接到互联网的人开始,然后向内缩小,而不是先从自己开始并向外扩展”;尝试精简启动方法,检验研究设计的可行性,如果研究项目注定失败,那么尽早失败,这样就不会浪费努力。

萨尔加尼克还讨论了在实际研究中计算社会科学遇到的伦理道德特定困难面^①及具体实用技巧^②。

四、计算模拟：被忽略的范式

计算社会科学“不同于以往借助社会调查抽样数据进行描述和经典模型回归分析的定量研究,而是借助复杂模型和社会计算工具对复杂社会现象与过程进行描述、解释和预测的定量社会学新领域”(萨尔加尼克,2019:311-315;Cioffi-Revilla, 2017: 1-7;陈云松,2020)。萨尔加尼克总结了数字时代的社会研究方法,给出了一般性原则。他强调,利用计算科学的现有技术,分析社交媒体等现有数据资源,是一个相对严格的计算社会科学范式(Cioffi-Revilla, 2017: 1-20)。因此,萨尔加尼克强调对数据的挖掘与利用,尤其是将研究者自己采集的数据与现成数据结合使用。虽然计算社会学在很多方面改变了传统社会学研究,但是萨尔加尼克的计算社会科学在本质上同传统社会研究仍是一致的,即基于实证的关联性视角,观察数据之间、探求变量之间的关系,依然是一种观察范式,其研究遵循演绎逻辑,只不过新技术拓展了演绎视角的研究范围。

但是,萨尔加尼克关于计算社会科学的讨论忽略了一个重要组成部分——计算模拟(simulation)方法(Cioffi-Revilla, 2017: 1-80)。社会模拟能够解决演绎视角下难以解释的社会问题。计算模拟通过计算机系统复制或者再现社会现象,“利用计算机编程灵活模拟行动者的行动与互动,观察可能产生的一系列社会后果,更好地理解人类经济与社会活动中涌现出来的模式及动态”(乔天宇、邱泽奇,2020)。社会模拟是一种生成性(generative)的社会科学研究范式(Epstein & Axtell, 1996: 1-20),即研究由相互影响的、自适应的行动者(agent)所组成的系统,这些行动者在物理空间或社会结构中占据位置,接受社会环境输入的信息,基于信息与规则展开互动,自下而上涌现出系统层面的结果,如合作等(Cederman, 2005; Epstein, 2006: 44-49),与传统社会研究中通常所见的观察范式有很大不同。因此,计算模拟是一种自下而上的研究路径,强调模型假设得

① 知情同意、理解与管理信息化风险、隐私权以及面对不确定性所做出的决策。

② 机构审查委员会是底线,不是上限;换位思考;将研究伦理视作连续的而非离散的过程。

到结果的普遍性。社会模拟模型可分为不同种类,其中最重要的是根据建模方式将模型分为基于变量的社会模拟和基于研究对象的社会模拟(Cioffi-Revilla, 2017: 1-80;乔天宇、邱泽奇,2020)。

社会模拟研究对于社会科学具有重要意义。首先,社会模拟使社会研究具有更大可行性。社会现象是复杂的,具有高维性、非线性与随机性等特征,观察范式往往无法处理这种高维数据。为了便于处理数据,经常需要假设行动者完全无知,只能进行随机搜索,或者是假设行动者全知全能,不需要探索就能做出决策,这些假设与现实数据存在很大距离。社会模拟则考虑到行动者与社会的多样性,可以设定行动者具有不同的行为规则与策略,并且会产生路径依赖、自组织以及噪音等结果,扩大现有模型的适用范围,使得研究结果更具普遍性(De Marchi & Page, 2014)。

其次,观察范式很难解释与预测黑天鹅、灰犀牛等事件。当前社会的连通性、不同个体与部分之间的互动、涌现出的稀少事件会对整个世界产生巨大影响,然而往往缺乏稀少事件的数据,制约了对稀少事件的研究。社会模拟方法通常可以得到复杂的结果,通过时间规模分布等途径,适合于研究极端的事件,提高社会科学的解释能力。

最后,社会模拟能够很好地应对数字时代计算社会科学所面临的伦理道德问题。许多社会研究由于伦理道德等原因,无法在现实社会中进行。社会模拟通常是一个计算机程序,现实中无法实现的研究往往可以在社会模拟系统中进行。基于观测数据与以往研究发现,设定不同假设与行动者策略,可以对社会线性产生过程与机制进行研究(Cioffi-Revilla, 2017: 1-20)。

总之,社会模拟研究具有更好理解复杂世界的潜力,社会研究可以模拟建构更加丰富、多样的行为与社会环境,作为演绎逻辑社会研究方法的重要补充,一起推动社会科学的发展。

五、总结

数字时代不断产生的痕迹数据,为研究者提供了记录与分析社会现象的机遇与挑战。萨尔加尼克认为当前时代对计算社会科学有许多误解,他希望研究者正视数字社会的趋势,用面向未来的眼光看待刚刚出现的计算社会科学。

随着计算方法进入社会科学研究领域,社会科学正在发生巨大变化。在传统调查、访谈与观察收集“小”数据的时候,社会科学研究者基于模型与理论假设进行统计推论,往往只是对已有研究的检验,而不是发现得出的观点;然而计算社会科学弱化对于理论模型的需要,将更多精力与创造力放在算法与数据上,不仅可以用数据来检验理论假设,还能从数据中发现更多新的内容。另外,计算社会科学的发展也拓展了因果解释可能性,开始强调传统社会科学研究所忽视的预测功能。因此,模型的分析策略也从传统社会研究的偏差最小,发展为减少模型的方差(Evans, 2020)。计算社会科学预测未来发生的事情以及确定影响因素,从而可以为社会政策等的制定提供更多支持,更好地服务于社会现实。

计算社会科学范式也代表了学术群体与社会世界之间新的关系。数字技术的一个重要影响就是改变群体之间的关系,在社会研究中的体现就是学术共同体与政府、企业等社会关系的变化。在传统社会研究中,研究者往往主导数据的收集过程(从研究设计,到数据收集);但是随着信息技术的兴起,无处不在的数字设备以及网络社会的连通性重新建构了社会结构,使得社会研究者对数据的主导权遭到挑战。数字设备可以在人们无察觉的情况下产生“高通量”(high throughput)数据,其获取与存储需要大量的资金与技术支持,企业、政府等组织是这些数据的重要拥有主体。研究者要想利用这些数据,就需要与企业或政府合作,失去了对数据的控制权。

数字技术不仅改变了学术群体与社会世界之间的关系,也改变了学术群体内部的关系。传统社会科学研究者基本来自社会科学领域,随着计算社会科学的出现,计算机等自然科学领域的研究者也开始进入社会研究领域。首先,大数据产生了对分析工具的需要,需要将非结构化的数字信息转变为可以分析的形式。许多计算科学家发明了处理海量数据的模型与分析工具,不仅可以处理文本,还可以处理图片、影像等信息。其次,数据革命带来社会分析革命。计算机等领域的学者通过 LASSO 等分析模型实现了对高维数据的降维分析,用随机森林、深度学习处理数据中的非线性交互作用解释或预测复杂社会现象(Evans, 2020)。因此,正如萨尔加尼克所说,计算社会科学需要促进社会科学家与数据科学家的合作,实现跨地域协作、跨方向交流,甚至是大规模协作等新的科研实践方式(陈云松, 2020)。

当前国际学界,计算社会学正在快速发展。对于我国社会科学来说,抓住计

算社会科学发展的机遇,是跟上甚至引领国际前沿研究领域的重要契机,对社会科学理论建设、新时代社会主义建设的政策支撑都有重要意义(陈云松、吴晓刚、胡安宁等,2020;罗玮、罗教讲,2015)。在新的社会研究范式形成过程中,中国社会科学界要认识到计算社会学的潜力,拥抱数字时代的影响。为此,需要做到以下几点:

首先,社会科学要以开放的心态接受计算社会学方法(陈云松,2020)。社会科学研究者已经意识到尚处在早期发展阶段的计算社会学存在的问题,但要用乐观积极的态度去推动其更好地发展。这要求社会科学研究者学习和了解计算社会学的范式与研究方法,具备计算社会科学的研究思维。只有这样,才能更好地适应计算社会科学的发展,同时对数字社会有更加深刻的认识。

其次,推动数据的公开与共享。数据是计算社会学发展的重要基础,但是大数据具有难以获取的特征,这制约着计算社会科学的发展。再加上当前我国在数据公开和共享方面发展还比较落后,数据获取的问题显得更加突出。另外,社会科学研究者需要尝试将现有数据与新的大数据资源结合起来,拓宽社会学研究的范围。

最后,关注数字时代的研究伦理:一方面,数字时代的研究者对个体生活的控制与干预能力有了极大提高,因此需要中国社会学界形成数字时代的研究规范和道德准则;另一方面,需要关注与讨论数字时代的隐私等道德问题,引导科技向善。

参考文献

- 陈云松,2020,《中国计算社会学的发展——特征、优势与展望》,《湖南师范大学社会科学学报》第5期。
- 陈云松、吴晓刚、胡安宁等,2020,《社会预测:基于机器学习的研究新范式》,《社会学研究》第3期。
- 罗玮、罗教讲,2015,《新计算社会学——大数据时代的社会学研究》,《社会学研究》第3期。
- 彭特南,2015,《智慧社会:大数据与社会物理学》,汪小帆、汪容译,杭州:浙江人民出版社。
- 彭玉生,2011,《社会科学中的因果分析》,《社会学研究》第3期。
- 乔天宇、邱泽奇,2020,《复杂性研究与拓展社会学边界的机会》,《社会学研究》第2期。
- 邱泽奇,2018,《大数据给社会学研究带来了什么挑战?》,《实证社会科学》第6卷。

- 邱泽奇、范志英、张樹沁, 2015, 《回到连通性——社会网络研究的历史转向》, 《社会发展研究》2015 年第 3 期。
- 萨尔加尼克, 2019, 《计算社会学: 数据时代的社会研究》, 赵红梅、赵婷译, 北京: 中信出版集团。
- 舍恩伯格、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。
- 宋庆宇、张樹沁, 2019, 《身体的数据化——可穿戴设备与身体管理》, 《中国青年研究》第 12 期。
- Ansolabehere, S. & E. Hersh 2012, “Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate.” *Political Analysis* 20(4).
- Banerjee, A. & E. Duflo 2009, “The Experimental Approach to Development Economics.” *Annual Review of Economics* 1(1).
- Barbaro, M., T. Zeller & S. Hansell 2006, “A Face Is Exposed for AOL Searcher No. 4417749.” *New York Times* 9(2008).
- Bohannon, J. 2016, “Mechanical Turk Upends Social Sciences.” *Science* 352(6291).
- Cederman, L. 2005, “Computational Models of Social Forms: Advancing Generative Process Theory.” *American Journal of Sociology* 110(4).
- Cioffi-Revilla 2017, *Introduction to Computational Social Science: Principles and Applications*, London: Springer.
- De Marchi, S. & S. Page 2014, “Agent-Based Models.” *Annual Review of Political Science* 17.
- Deaton, A. 2010, “Instruments, Randomization, and Learning about Development.” *Journal of Economic Literature* 48(2).
- Dickinson, J., B. Zuckerberg & D. Bonter 2010, “Citizen Science as An Ecological Research Tool: Challenges and Benefits.” *Annual Review of Ecology, Evolution, and Systematics* 41.
- Epstein, J. & R. Axtell 1996, *Growing Artificial Societies: Social Science from the Bottom up*. Washington: Brookings Institution Press.
- Epstein, J. 2006, *Generative Social Science: Studies in Agent-Based Computational Modeling*, New Jersey: Princeton University Press.
- Evans, J. 2020, “Social Computing Unhinged.” *Journal of Social Computing* 1(1).
- Farber, H. 2015, “Why You Can’t Find A Taxi in the Rain and other Labor Supply Lessons from Cab Drivers.” *The Quarterly Journal of Economics* 130(4).
- Ginsberg J., M. Mohebbi & R. Patel et al. 2009, “Detecting Influenza Epidemics Using Search

- Engine Query Data.” *Nature* 457(7232).
- Marshall, P. , C. Lintott & L. Fletcher 2015, “Ideas for Citizen Science in Astronomy.” *Annual Review of Astronomy and Astrophysics* 53.
- Mas A. & E. Moretti 2009, “Peers at Work.” *American Economic Review* 99(1).
- Narayanan A. , J. Huey & E. Felten 2016. “A Precautionary Approach to Big Data Privacy.” In S. Gutwirth, R. Leenes & De Hert eds. , *Data Protection on the Move*, Dordrecht: Springer.
- National Research Council 2013. *Nonresponse in Social Science Surveys: A Research Agenda*. [S. I.] : National Academies Press.
- Restivo, M. & A. van de Rijt 2014, “No Praise Without Effort: Experimental Evidence on How Rewards Affect Wikipedia’s Contributor Community.” *Information, Communication & Society* 17(4).
- Salganik, M. , P. Dodds & D. Watts 2006, “Experimental Study of Inequality and Unpredictability in An Artificial Cultural Market.” *Science* 311(5762).
- Schober, M. , F. Conrad & C. Antoun et al. 2015, “Precision and Disclosure in Text and Voice Interviews on Smartphones.” *PloS One* 10(6).
- Schultz, P. , J. Nolan & R. Cialdini et al. 2007, “The Constructive, Destructive, and Reconstructive Power of Social Norms.” *Psychological Science* 18(5).
- Seltzer, W. & M. Anderson 2008, “Using Population Data Systems to Target Vulnerable Population Subgroups and Individuals: Issues and Incidents.” In J. Asher, D. Banks & F. Scheuren eds. , *Statistical Methods for Human Rights*, New York: Springer.
- Watts, D. & S. Strogatz 1998, “Collective Dynamics of ‘Small-World’ Networks.” *Nature* 393(6684).
- Watts, D. 2014, “Common Sense and Sociological Explanations.” *American Journal of Sociology* 120(2).
- Zimmer, M. 2010, “But the Data Is Already Public: On the Ethics of Research in Facebook.” *Ethics and Information Technology* 12(4).