

数字社会与计算社会学的演进

邱泽奇

内容提要 过去 40 年是数字社会发展的 40 年。数字技术的社会化应用改变了社会经济基本动力,形成了与工业社会具有本质区别的数字社会,并使数字社会进入了飞跃发展的阶段。数字社会与工业社会的本质差异至少可以从个体与社会的关系、社会分化机制两个维度管中窥豹。然而,社会学研究在方法上未能有效跟上数字社会发展的时代步伐。虽然有人在 20 世纪 90 年代就提出了计算社会学的概念,但直到大约 10 年前,社会学研究方法的突破还依然停留在概念上。今天,ABM 的应用和数据挖掘技术的发展为社会学研究方法的突破带来了转机,形成了计算社会学发展的两条演进路径,其应用正处于相向而行的态势之中。

关键词 数字社会 计算社会学 ABM 数据挖掘

邱泽奇,北京大学博雅特聘教授,教育部“长江学者”特聘教授,北京大学中国社会与发展研究中心主任

自从大数据(big data)概念从商业应用向学术研究传播以来,社会学界对数字技术变革影响社会变迁的反应也随即展开,其中的一个反应是计算社会学的兴起。不过,计算社会学的发展并非线性式的一帆风顺,而是与数字社会的发展相伴随。一个直观的事实是,小布伦特早在 1993 年便使用了计算社会学(computational sociology)概念,并认为计算社会学是社会学下一个千年的希望^[1]。可是,小布伦特的文章没有发表在社会学期刊上,而是发表在了计算机类期刊上。这篇文章发表之后,无论是计算社会学概念还是小布伦特的文章都没有在社会学界引起任何反响,直到当下为止,谷歌统计的文章被引次数也只有 6 次。众所周知的事实还有,一段时间内,计算社会学无论是作为社会学的一种方法,还是作为社会学的子学科,都不仅没有像小布伦特倡导的那样蓬勃发展,甚至没有引起主流社会学家们的足够重视。直到近 10 年来,随着数字社会的极速发展,运用大数据观察和探讨社会发展变迁的计算社会学才凸生显现,并吸引了人们越来越多的关注。

[1]Brent, Edward E., "Computational Sociology: Reinventing Sociology for the Next Millennium", *Social Science Computer Review*, 1993, 11(4), pp.487-499.

假设计算社会学概念出现在20世纪90年代初期^[1],那么值得探究的是,一个有力且正确的洞察为什么在沉寂了近30年之后才形成学术潮流?本文试图把计算社会学的演进放在数字社会发展环境里,探讨数字社会发展带给计算社会学的机会以及计算社会学的演进过程。

一、数字社会的发展

用社会学期刊文章的眼光看,小布伦特的文章可以被认为是一篇糖水文章,嘟囔着社会学的经典八卦,如定性与定量、理论与方法、微观与宏观等等,既没有独到的证据链,也没有经过严密论证的学术观点。可是,我们却不能因此否认文章的学术洞察力和洞察的正确性。他从计算机算力(computational power)和应用软件的发展出发,与总体社会复杂性研究建模与计算能力的需要类比,推演出计算社会学概念,倡导社会学家们应该想得大一些,合作规模大一些,共同探索和把握之前的社会学家们没有机会探讨的社会现象和规律。非常遗憾的是,小布伦特限于想象力,没有告诉社会学同行们计算社会学到底是一种怎样的学问,可以怎样去探索和发展。一句话,小布伦特从计算机科学的发展直观地联想到社会学进行大规模建模和计算的需要,认为大规模计算应该能发现非常不一样的社会规律,它将给社会学下一阶段的发展提供机会。

现在看来,小布伦特的洞察和倡导是正确的。可是,他的倡导发出之后,为什么没有很快获得社会学界的反馈?在我们看来,最直接的影响因素是数字社会发展的阶段性,即20世纪90年代初期是数字社会发展的初始期。计算机的商业应用虽早在20世纪60年代已经开始,可直到20世纪90年代还处在机构性组织应用阶段,不仅没有产生巨量数据,仅有的数据社会学家们也难以接触到。

我们知道,计算机只是数字社会发展的必要条件,充分条件是互连网络。计算机网络虽在20世纪60年代就已经发明,却一直限于军事和科研用途,直到1990年才出现向社会扩散的第一个充分条件,即网页浏览器的发明^[2]。1993年图形界面万花筒(Mosaic)浏览器的出现为互连网络的社会化应用提供了大众可以使用的工具。可是,仅有浏览器,没有网络设施和设备接入服务,浏览器也只是一个没有意义的单机玩具。1994年互连网接入商业服务的出现为数字社会的发展集齐了基本条件,即终端(计算机)、网络和浏览器(网络人机界面)等三驾马车,让行动者有了实现互联互通的机会,互连网的社会化应用才真正出现了^[3]。中国科学院高能物理研究所1993年开通了与美国西海岸的第一条专线,标志着中国正式接入互连网世界,1997年北京瀛海威有限公司提供互连网接入服务,标志着互连网在中国进入社会化应用,数字社会正式进入其发展轨道。

在技术扩散和流行现象研究领域,有一条著名的S曲线,它把一项技术、潮流或时尚的流行划分为五个阶段。以目标人群为分母,以采用或使用人群为分子,计算采用某项新技术或进入某个潮流或时尚人群占目标人群的比例,用这个占比划分发展阶段。第一阶段为前2.5%阶段,即先行种子阶段。在这个阶段,只有愿意尝新的约2.5%者会一试。第二阶段为前13.5%阶段,即前流行阶段。在这个阶段,有约13.5%者会采用或使用,形成流行前的先锋征兆。第三阶段为前34%阶段,即流行拐点阶段。这个阶段,已有1/3者会采用或使用,形成因网络效应产生三度影响进而让扩散加速的格局^[4]。跨过这个拐点,如果没有其他干扰,流行将变得不可避免。第四个阶段为后34%阶段,即后流行拐点

[1]小布伦特的文章是我们查到的最早提出计算社会学概念的文献。由于没有穷尽文献,故以假设论。

[2]蒂姆·伯纳斯-李在1990年发明了第一个网页浏览器World Wide Web,此浏览器后改名为Nexus。

[3]最早接入互连网的社会性应用出现在美国麻省的莱星顿(Lexington)和剑桥(Cambridge)社区。

[4][美]尼古拉斯·克里斯塔基斯、詹姆斯·富勒:《大连接:社会网络是如何形成的以及对人类现实行为的影响》,简学译,中国人民大学出版社2013年版,第39页。

阶段。这个阶段已有2/3者会采用或使用,形成了流行的初步饱和。此后,流行的速率明显降低。第五阶段为后16%阶段,即流行高原阶段,剩下的不一定会采用或使用^[1]。

由此我们说,过去的40年正是向数字社会的转型阶段。自工业革命以来,工业生产的革命性组织方式催生了金融市场,即对社会影响巨大的企业都在金融市场上^[2]。上市公司的市值排名不只标示着企业的价值规模,也指示技术和行业发展的潮流和社会的取舍,上市公司市值世界前十的排序,对理解主导性社会力量具有风向标意义。根据公开数据,在市值排名世界前十的公司中,1990年有6家银行,1家通信硬件公司;2000年有7家通信硬件公司,1家互联网公司;2010年有4家能源公司,2家互联网公司,2家银行;2020年有7家互联网公司,2家金融公司^[3]。过去40年,主导力量的转换显示了数字社会发展的线索。那就是,在21世纪之前,数字社会处于萌芽期。进入21世纪后,先是网络基础设施的发展,接着是终端设施设备的发展,然后是数字社会的真正来临,一个直接指标是,2021年脸书系社交应用总月度活跃的用户数达到34.5亿,占世界总人口的43.7%,早已迈过了数字技术扩散的拐点,意味着数字技术真正进入了人们的日常生产和生活。

中国的经历表明,以互联网络设施设备等硬件为基础,以组织和个体应用为目标,数字技术快速迈过其扩散拐点,进入数字社会的腾飞阶段^[4]。在经济领域,2020年数字经济增加值为39.2万亿元人民币,占GDP的比重跃升至38.6%,一下子越过了技术扩散起飞的临界点,进入数字经济的腾飞期;预计2025年数字经济占GDP的比重将超过50%^[5],中国真正进入数字经济时代。

在社会领域,中国上网人口从1993年的153人极速上升至2020年近9亿人。其中,有99%使用移动终端上网。在不到十年时间里,中国完成了从传统互联网向移动互联网的演化。到2020年,中国已拥有世界上规模最大的光纤网络、数量最多的4G和5G基站、覆盖人口最广泛的数字技术用户,电子商务应用覆盖了最不可能覆盖的所有832个前国家级贫困县,数字教育设施设备覆盖了所有中小学,包括最难覆盖的偏远地区自然村寨教学点。从发展进程看,在刚刚过去的“十三五”期间,中国上网人口增加了43.7%,是上网人口增加最集中的时期。从人群结构看,在上网总人口中,60岁以上上网人口的占比比60岁以上人口占总人口的比重高出11.0个百分点;城乡上网人口占比差距正在快速缩小^[6]。

在政务领域,从建站到联网,省部级政务接入国家政务平台的比例快速上升,居民跨行政区办理业务的便利性获得极大增强。在政务平台上,实名注册个人数达7.74亿,法人达7.27亿。审批等行政许可事项实现网上受理和最多跑一次的比例达82.13%;一半以上行政许可事项平均时限压缩超过40%;政府事项网上可办率超过了90%^[7]。不仅如此,政府作为机构还直接参与“社交”,几乎所有社交平台,都有政府号。数字连接正推动中国治理体系和治理能力的现代化。

根据国际电联(ITU)的数据,以互联网个人用户为统计口径,世界上尽管还有近一半人口没接入

[1]关于S曲线,参见〔美〕埃弗雷特·M.罗杰斯:《创新的扩散》,辛欣译,中央编译出版社2002年版,第244—245页。

[2]〔日〕板谷敏彦:《世界金融史:泡沫、战争与股票市场》,王宇新译,机械工业出版社2018年版,第125—130页。〔奥〕卡瑞恩·克诺尔·塞蒂娜、〔英〕亚历克斯·普瑞达主编:《牛津金融社会学手册》,艾云等译,社会科学文献出版社2019年版,第37—38页。

[3]依据公开数据整理,2020年的数据来自普华永道(PwC)官方网站。系统的叙述,参见徐清源:《数字企业平台组织的结构、行动和治理》,北京大学博士论文,2021年。

[4]〔6〕〔7〕主要综合自中国信息通讯研究院:《中国数字经济发展白皮书》,中国信通院,<http://www.caict.ac.cn/kxyi/qwfb/bps/>;中国互联网络信息中心:《中国互联网络发展状况统计报告》,中国互联网络信息中心,<http://www.cnnic.net.cn/hlwtzyj/hlwzxbg/>。在叙述中,对两类报告均参考了其初期和近两年的数据。

[5]在2021年的博鳌亚洲论坛上,经济学家李稻葵和国家网信办做出了相似的预测。

互联网,可是从数字技术扩散视角出发,数字社会早已跨过了其进入腾飞阶段的扩散拐点,并把人们带进了一个崭新的数字时代。

二、数字社会带来的挑战

随着世界迈进高度互联的数字社会,人类行为生产痕迹数据的来源大大增加,从计算机和手机使用,包括通信、文档交换、生产、交易、教育、健康、医疗等,到物联网关联的各类传感器,产生数据的设施设备呈几何级数增加,网设施设备的数量早已远远大于人类人口的数量。数字设施设备数量的极速增长,带来的结果之一是数字数据量呈几何级数的快速增长。2010年人类积累的数据大约2ZB,远远超过了人类过去几千年积累的数据。2020年则达到了64.2ZB,10年间,人类积累的数据量便增加了30多倍。预计到2025年会达到181ZB^[1]。

与传统社会学研究使用的文献数据、访谈数据、抽样调查数据、统计数据等容量和数量相比,数字社会可以用于社会研究的数据量已是一个天文数字,是社会学家们凭借过去近百年积累的经验和技术无力运用的,也因此对社会学研究的方法提出了尖锐的挑战。不仅如此,人类对数字技术和巨量数据的各类应用^[2]正深刻地改变着社会,从个体与社会的关系,到社会设置、经济生活、社会生活、政治环境等,再到社会分化的底层逻辑,都在发生深刻且本质的变化。

以社会学知识中的个体与社会关系为例。在数字社会之前,从发生学视角观察,个体-社会关系的缔结是经由已经存在的行动者实现的,即以人为媒。从生物性出生到社会性出生,从意识塑造、行为内化、文化内化到价值观内化,个体-社会关系的建构和发展都以人为媒。社会学的概念“社会化”试图归纳的便是个体-社会关系建构的复杂社会过程。在个体成长中,活动空间的地方性使得个体从社会化开始便是属地的,便是嵌入已经存在的社会势力如派系之中的,也因此被动地获得着自己的社会地位或身份,由此,形成了在地方性社会中个体及其与社会关系特征的公开性,并随时间的延续还塑造了关系的系统性。正如涂尔干反复强调的,社会是外在于个体、对个体具有强制性的力量^[3]。的确,个体的身体和情感等社会特征来自地方性社会。在个体真正成为一般意义的社会行动者后,又将其还给了社会。

然而,数字社会从本质上改变了个体-社会的关系缔结方式。数字连接的实现让个体-社会关系的建构可以跳出地方性社会,进而摆脱以人为媒的过程。个体运用数字连接可以直接建构与社会的关系。作为生物性个体,人们依然出生在地方性社会,依然在地方性社会成长,依然与家庭、身边的同伴发生连接。与此同时,作为社会性个体,从有能力使用数字工具开始,人们便有机会与地方性社会之外的广大社会建立联系,建构属于个体自己的“社会”,而不再是被动地接受地方性社会强加的关系。当然,以数为媒的社会与以人为媒的社会之间不存在一条界限清晰的鸿沟,而是彼此交集地混杂在一起。尽管如此,对个体-社会关系的建构而言,从社会意识塑造、行为内化、文化内化到价值观内化,个体-社会关系的建构和发展,都有了另一种完全不同的选择,即以数为媒。以数为媒赋予了个体一种本质上不同的机会,个体可以是自己的,同时又是世界的,而不仅仅是地方的;个体依然嵌入在已经存在的社会中,可这个社会也不再是地方的,而可以是部分身份或资格的、场景性的、离散的。在数

[1]IDC, Statista, "Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2025 (in Zettabytes)", Statista, Statista Inc., 7 Jun 2021, <https://www-statista-com.ezproxy.gavilan.edu/statistics/871513/worldwide-data-created/>.

[2]如生产、生活、政务、管理等各领域、各方向的应用。

[3][法]涂尔干:《职业伦理与公民道德》,渠敬东译,商务印书馆2017年版,第66—67页。

字社会,个体可以只把自己的数字属性交给社会而保留自己的实体属性,当然也可以两者都交给社会。只是,这个社会绝不再只限于地方性社会。

如果说以人为媒的个体-社会关系中的个体是地方性社会系统的,那么,数字社会叠加的以数为媒的则是世界性社会系统的。个体可以不再受到地方性社会的约束而拥有一个自己的、世界性的舞台,形成社会行动者的泛在连接。

与此同时,数字社会改变的不只是个体-社会关系,而是整个因工业化、市场化、城市化建构的社会体系,包括由政治、经济、社会关系体系形塑的社会分化机制。如果说,数字连接在个体-社会这一元关系(metaconnection)上改变了社会的底层组织逻辑,那么,泛在连接作为自变量给社会整体带来的机制性影响也是社会学关注的最重要变量,或可以归纳为社会分化的革命性变革。

在 20 世纪下半叶的社会学研究中,社会学家们乐此不疲的努力是沿着布劳-邓肯模型^[1],在主权国家范围内讨论汇集在个体身上的因素主导的社会分化。殊不知,在这类研究中有两个默认的前提假设是值得怀疑的(因为在其他社会不一定存在)。第一,给定制度。理论与实证研究都证明,制度是影响社会分化的主导变量,只有在制度稳定的前提下,其他因素的影响力才有机会释放^[2]。像美国那样制度相对稳定的社会在世界上并不是主流。第二,给定技术。理论与实证研究同样证明,在给定制度的前提下,技术变革也是影响社会分化的主导力量,只有在技术变革相对稳定且连续的前提下,其他因素的影响力才会对社会分化产生有效影响^[3]。可是,即使像美国那样制度相对稳定技术变革也是活跃的社会,也未必如此。《财富》杂志对上市公司市值排名的历史数据显示,前十公司的名单每十年都有本质变化,正如,前文述及过去 40 年市值排名世界前十公司名单的变化也证明了没有“常青藤”公司,其根本影响因素正是技术变革。

一旦我们从美国社会学“主流”跳出来,向布罗代尔^[4]学习,甚至超越布罗代尔,从更加长远的历史观察社会分化的事实,就会发现,人类正是因元关系地理范围的不同,才有不同的社会分化机制。以最简约方式归纳这个不同,我们可以将整个人类的历史划分为三个阶段:家庭化连接的散点社会,职业化连接的区块社会,以及数字化连接的个体社会。

在散点社会,如分散的乡土社会,社会分化以家庭为单位。分化的结果是以家庭为单位的、在极小地理空间的经济社会地位结构,最终形成的是由外在权力赋予或因经济力量获得的家庭政治地位。一个家庭在当地拥有的权势大小取决于一个极小的相对独立地理空间的以权势衡量的社会结构。在不同空间之间,家庭的权势不具有通约性,甚至不具有可比性。

在区块社会,如相对独立却又有可能连接的城市社会,社会分化以个体为单位,家庭经济社会地位对个体有影响,却不像散点社会那样直接赋予个体身份,而只能为个体提供基础和便利,这也是布劳-邓肯学术潮流纳入家庭经济社会地位变量的理论基础。在给定制度的前提下,个体经济社会地位的获得依靠其职业在社会分工体系的地位获得第一次机会,再通过受教育程度和技能的适配性获得第二次机会。与散点社会的社会分化不同,在区块社会的社会分化中,政治地位分化与经济地位分化不再紧紧相连,而各有自己的路径。当然,在不同的制度体系下,政治地位与经济地位的分离程度也

[1]Blau, P. M., Duncan, O. D., *The American Occupational Structure*, New York: John Wiley and Sons, 1967.

[2]邱泽奇、刘世定:《社会板块结构的变迁》,《中国社会科学》(内部文稿)2013 年第 6 期。

[3]Kohler, Timothy, A. et al., "Greater Post- Neolithic Wealth Disparities in Eurasia than in North America and Mesoamerica", *Nature*, 2017, 551(7682), pp.619-622.

[4][法]费尔南·布罗代尔:《日常生活的结构:可能和不可能,十五至十八世纪的物质文明、经济和资本主义》,顾良、施康强译,商务印书馆 2017 年版,第 1—3 页。

不相同。在经历了资产阶级革命的制度中,旋转门制度把两者联系在了一起,不过,这一联系,也仅限于精英阶层。在没有经历资产阶级革命的制度中,政治地位与经济地位至少在制度上不具有可兑换性。分化的结果是以个体为单位的、在更大地理空间如城市的个体社会经济地位结构,最终形成的是由行业相对地位赋予的个体经济地位,个体在区块空间的经济地位高低取决于在一个与其他区块有关联却又相对独立的地理空间的以收入和财富衡量的社会结构。一个典型隐喻是,个体手握相同数量财富居住在不同城市便具有不同的社会经济地位。

在个体社会,如当下正在发展的数字社会,社会分化依然以个体为单位,家庭社会经济地位对个体依然有影响,可相比于区块社会,如果个体希望沿袭区块社会的分化路径,其影响路径依旧存在。除此之外,个体还有新的路径,即泛在连接带来的让个体绕过政治地位和经济地位而直接进入社会分化或获得社会地位的路径。如果说散点社会的社会分化路径是政治路径,在政治、经济、社会三角中政治地位具有主导性,区块社会的社会分化路径是经济路径,在政治、经济、社会三角中经济地位具有主导性;那么,个体社会的社会分化路径则是社会路径,在政治、经济、社会三角中社会地位具有主导性。如果政治地位是权力的大小,经济地位是资本的多少,那么,社会地位便呈现为流量的巨细。在给定制度环境里,个体可以不再依靠家庭社会经济地位、行业产业地位,而依靠个体特征获得流量进而获得社会地位,流量甚至可以穿透劳动分工体系、穿透主权国家体系,成为人类社会共识的社会地位标准。分化的结果是,个体社会地位的高低取决于以个体为单位的、理论上可以在世界范围内至少在连接范围内的以流量衡量的社会地位结构,最终形成的是由圈子赋予的个体社会地位,即在一个圈子社会结构里的社会地位。

散点社会和区块社会的社会分化形成的是一个地理空间的整体性社会结构,个体社会的社会分化形成的则是平行的、跨域地理空间的部分性社会结构。圈子性社会结构是数字社会与之前的社会结构的本质区别。

值得进一步说明的是,从散点社会到个体社会,支撑社会分化机制变化的还有个体生活保障安全的逐步增强。如果说在散点社会和区块社会的社会分化底层逻辑中还有人类生存威胁带来的影响,即家庭和个人努力争取政治地位和经济地位的动力来自对家庭或个体生计保障的考量,那么,在个体社会,个体争取社会地位可以被理解为是纯粹对自我特征社会性认同的努力。

数字社会的本质特征之一是个体社会的发展。个体化的发展和社会行动者之间的高度互联,加上普遍渗透的物联网传感器的使用,让人类社会在极短的时间内进入高度复杂性时代。让我们再次回到更加长远的历史来观察当下的数字社会。散点社会延续了万年之久,区块社会虽然发端久远,真正兴盛且对人类社会产生实质性影响也只是在第一次工业革命之后的不到300年。可是,如前述及,即使我们把浏览器发明的1990年当作数字社会元年,数字社会也不过短短40年时间。可是,即使只观察个体-社会关系和社会分化机制的变迁,数字社会也呈现出高度复杂性。

遗憾的是,社会学家们好像被数字社会的发展远远地甩在了后面而未特别觉醒。在经历了20世纪60—70年代的繁荣后,社会学在20世纪80年代进入沉寂期,在理论和研究方法的发展上均陷入停滞。在理论上,宏大理论叙事在被推倒之后,理论发展陷入停滞;在研究方法上,二战后繁荣的实证研究路径遇到了社会复杂性的挑战。社会学理论与研究方法都在寻找下一个发展的突破口。按常理,社会学家们应该为数字社会的发生与快速发展做好了准备,且积极拥抱数字社会的来临;可事实是,社会学家们并未预见数字技术带来的革命性社会变革,无论是对社会复杂性的几何级数跃升还是对计算机科学带来的算力高速发展,都视而不见。倒是未来学家们不断推测未来趋势,却也只把眼光停

留在当时美国面临的能源危机和持续经济增长乏力难题上^[1]。没有人敏锐感到一场真正社会革命的来临。

直到21世纪曙光初现,左翼社会学家卡斯特出版信息社会三部曲^[2],才给人们刻画了一个因互联网高度互联而改变的社会。遗憾的是,卡斯特只是刻画了普遍连接带来的影响,却没有洞察互联网社会的本质变化是因人类社会高度互联带来的数据积累与应用的革命性影响。我们认为,因数字技术而改变的社会不单纯是网络社会。网络只是发生了相对于散点社会和区块社会的社会连接形态改变。其实,社会网络是自社会诞生以来就有的社会连接形态,只是随着时代的发展,网络规模有大小之别而已,其本质上还是网络。无论是村寨的人际网络还是电信技术发明之后的电报网络和电话网络,都是网络。只有当网络连接数字化,社会才真正发生了本质改变,个体-社会关系的变化只是一个可以体验和想象的例子。因数字技术而改变的社会也不单纯是信息社会,信息社会归纳的是信息生产、分发、配置、整合等信息很活跃的社会^[3],信息社会可以不基于数字技术,也不对社会带来本质改变。只有当信息活动源于数字技术又回归数字技术,信息被埋藏于巨量数据之中,社会才发生了本质改变,数字社会分化机制的变化也只是本质改变的一部分。

简单地说,要理解数字社会的特征,在研究方法上,源起于工业社会的思辨和实证等研究路径还可以沿着既有路径前行,却已显然力有不逮。社会学学科的发展亟须在研究方法上进行突破以赶上数字社会的发展潮流,开发适用于巨量数据和高度复杂性分析的研究方法,以及基于新研究方法的研究策略和路径。

三、计算社会学的演进

社会学家们对社会复杂性的警觉自社会学诞生之初就已经存在,从涂尔干到布劳^[4],都认为社会学方法在应对社会复杂性上有困难。如在具体研究中区分微观社会学和宏观社会学只是应对复杂性不得已的选择。尽管社会学家们始终在探索应对社会复杂性的方法策略和技术。但遗憾的是,让社会学有能力应对社会复杂性的愿望始终未能实现。自小布伦特的文章之后,1995年才有人感到计算机技术扩散的影响,建构了计算社会学的“理论-经验-计算”三角模型^[5]。显然,这是从社会学统计方法衍生出来的模型,尚没有关注巨量数据给社会研究方法带来的影响和压力。

一晃又是7年,直到2002年社会学主流期刊终于刊登了第一篇以计算社会学为主题的文章。梅西和惠勒虽然没有明确指出却隐晦地告诉读者,传统社会学研究在方法上是“因素”(factors)的研究,即使在计算机出现后,社会学家们也只是用计算机提高因素关系的计算效率,包括仿真计算。基于数字技术的影响已经让社会结构和社会秩序超越了人居地理空间,基于行动者动态的研究需要引入一

[1]参见〔美〕阿尔温·托夫勒:《第三次浪潮》,朱志焱、潘琪译,生活·读书·新知三联书店1983年版,第130—132页、第376—377页。〔美〕丹尼尔·贝尔:《后工业社会的来临:对社会预测的一项探索》,高钰等译,商务印书馆1984年版,第509—512页。

[2]卡斯特的信息社会三部曲,参见〔美〕曼纽尔·卡斯特:《网络社会的崛起》(另两部为《认同的力量》《千年的终结》),夏铸九等译,社会科学文献出版社2001年版,第1—5页。

[3]Soll, Jacob, *The Information Master: Jean-Baptiste Colbert's Secret State Intelligence System*, Ann Arbor: The University of Michigan Press, 2009.

[4]乔天宇、邱泽奇:《复杂性研究与拓展社会学边界的机会》,《社会学研究》2020年第2期。

[5]Hummon, Norman P., Thomas J. Fararo, "The Emergence of Computational Sociology", *The Journal of Mathematical Sociology*, 1995, 20(2-3), pp.79-87.

种新方法。为此,他们介绍了基于行动者建模(agent-based modeling, ABM)的方法^[1]。这是自计算社会学概念提出以来第一个实质性进展,即计算社会学有了可用的方法工具。

或许是得益于文献的扩散效应,社会科学领域其他学科尤其是人类学、经济学、政治学、心理学等^[2]学科的计算方法转型也快速跟进,形成了一个新的概念:计算社会科学(computational social science, CSS)。2009年拉泽尔等人的文章预示了计算社会科学的诞生^[3]。10多年之后,拉泽尔等人再次撰文探讨计算社会科学的机会和发展遇到的障碍^[4],这也意味着计算社会科学的发展并非一帆风顺。

与梅西和惠勒的路径不同,拉泽尔等人从数字技术带来的巨量数据出发,提出了从数据中挖掘人类个体或群体社会行动模式的策略,即数据挖掘(data mining)策略。在社会科学中,数据挖掘不是一个新概念,早在1983年经济学家就提出了数据挖掘概念^[5],只不过那时这个概念还没有把巨量数据纳入视野。随着20世纪90年代数据量的积累,商业公司开始进入数据挖掘领域以获取商业利润,提出了数据库挖掘(database mining)概念。随后,研究者将数据库挖掘简化为数据挖掘,且出现了许多平行概念,如知识发现(knowledge discovering)、数据考古(data archaeology)等。随着1995年第一届数据挖掘与知识发现国际会议(KDD-95)的召开,数据挖掘正式进入学术界,成为运用巨量数据探讨人类社会规律的方法,也成为计算社会学发展的另一条路径。

简言之,从20世纪90年代初期提出计算社会学概念到21世纪初梅西和惠勒提出第一条方法路径,再到10年之后的拉泽尔等人提出的第二条方法路径,计算社会学完成了从概念到研究实践的工程工艺过程。从21世纪的第二个十年开始,计算社会学沿着这两条路径在快速发展。

在ABM领域,随着从单主体互动模型向多主体互动模型(ABMs)的发展,ABM在社会学领域的应用日渐扩展,方法也日臻成熟。斯卡佐尼归纳了ABM在社会学的应用场景和方法路径,从模型复证和多层效度入手,形成了系统的方法实践逻辑和知识。由此ABM路径也被引入国内,并在社会科学领域获得了广泛的应用,正如在社会学领域应用于对交通出行、社交行为、社会信任、不平等效应、公共政策的评估等^[6]。吕鹏等人还基于群体智能、社会动力模型以及多主体仿真等方法,探讨了在面临群体性共同危险(如恐怖袭击)情况下,个体见义勇为、勇斗歹徒(如英雄行为)的人群动力学机制^[7]。为探讨ABM模型的多用性,我们还尝试用ABM检验实证研究的结论^[8]。

乡村电商是过去10年中国乡村发展的一大景观。2009年阿里巴巴发现了3个淘宝村^[9],2020年淘宝村的数量增长到5425个,加上运用其他电商平台从事电商经营的电商村,中国的电商村预估超过万个,总数量超过了中国行政村总数的2%,覆盖了大多数前国家级和省级贫困县和贫困村。我们知

[1]Macy, Michael W., Robert Willer, "From Factors to Actors: Computational Sociology and Agent-Based Modeling", *Annual Review of Sociology*, 2002, 28(1), pp.143-166.

[2]加上社会学,在社会科学领域通常也被称之为社会学科学的“五大学科”(big five)。

[3]Lazer, David, et al., "Computational Social Science", *Science*, 2009, 323(5915), pp.721-723. 计算社会科学的发展可参见张小劲、孟天广:《论计算社会科学的缘起、发展与创新范式》,《理论探索》2017年第6期。

[4]Lazer, David, et al., "Computational Social Science: Obstacles and Opportunities", *Science*, 2020, 369(6507), pp.1060-1062.

[5]Lovell, Michael C., "Data Mining", *The Review of Economics and Statistics*, 1983, 65 (1), pp.1-12.

[6]相关文献通过中国知网的主题词检索和关键词检索均可获得,为节省篇幅,恕不一一列出。

[7]Peng Lu, et al., "Swarm Intelligence, Social Force and Multi-Agent Modeling of Heroic Altruism Behaviors under Collective Risks", *Knowledge-Based Systems*, 2021(214), <https://doi.org/10.1016/j.knosys.2020.106725>.

[8]邱泽奇、黄诗曼:《熟人社会、外部市场和乡村电商创业的模仿与创新》,《社会学研究》2021年第4期。

[9]淘宝村是阿里研究院发明的概念,可参见阿里研究院网站。

道,电商经营是有数字素养门槛的新技术,受教育程度不高的村民又是如何了解并掌握新技术、运用新技术促进乡村发展的呢?我们从实地调查中获得的结论是,一些外出打工的年轻人在城市里学得运用计算机、互联网甚或电商经营技术,回到村里在线上销售村里的产品,形成了乡村电商的星星之火。乡村的熟人社会特征让任何一种赚钱的门道都有可能在校内经由亲属和邻里网络形成社会学习氛围而快速传播,让任何一个想学习的村民都有机会学习、模仿和创新,由此构成了电商技术的社区扩散。其中,在给定外部市场无限的前提下,有人带回电商技术且经营电商是必要条件,有人愿意学习、模仿和创新是充分条件。

可是,关于星星之火究竟如何变成燎原之势的,我们只能听村民讲过去的故事。由于社会学传统研究方法无法复证村民的故事,因而无法将村民故事中抽象的规律变成可复证的知识。运用ABM仿真,我们不仅重现了村民的故事,还发现了新的知识,那就是村民网络密度和关系强度对社会学习有效性的潜在且具实质性的影响,即两者的影响随村民学习能力的增强而被放大。当村民吸收异质能力、模仿与创新潜力、创造能力都处在中高强度时,三者的互补性效应不仅创造了显著的个体收益和群体收益,且整体收益分配也没有出现典型的幂律分布(强者愈强、弱者恒弱),而是出现了“涓滴效应”,即先富带后富。随着网络密度的减小或整体关系强度减弱,收益创造和收益分配的两个社会学习效益都明显减弱;当网络特征趋近于陌生人社会时,即便给定相同的能力设置,收益创造和收益分配的两个社会学习效益也趋于消失,即三类学习能力的互补效用无法发挥。

对比仅利用乡村调查数据的实证研究,ABM仿真在检验实证研究结果的基础上还创造了一项对比实验,用于观察和检验熟人社会作为一种环境对电商技术扩散的影响,反向证明了村民社会网络对电商技术扩散影响的重要性。

在数据挖掘领域,随着机器学习技术的发展,研究的社会现象范围越来越宽,内容也越来越深。不过,总体上依然以信息提取、社会网络拟合、社会复杂性拟合和社会仿真为基本路径^[1]。在应用中,从数据出发的研究路径几乎扩展到了所有具有巨量数据的领域,形成了相对模式化的方法流程:数据挖掘三步法。第一步,面对杂乱无章的非结构化巨量数据,运用数据样本进行标注,无论采用直接计算方法还是采用机器学习方法,形成数据图谱(map)。第二步,面对数据图谱,通过聚类合并或其他方法合并和剔除非主题关联数据集(reduce),形成研究数据集。第三步,运用关联数据集进行模型训练,用于知识发现,并将获得的模型放回流动的数据中进行迭代优化。经典的例子如谷歌运用用户搜索数据建构的流感模型^[2]。

在社会学研究中,数据挖掘已广泛运用于医疗、教育、家政、养老、出行等社会生活领域,也广泛应用于社区安全、社会服务、社会冲突处理等社会治理领域。经典的例子如约翰逊等运用社交数据对伊斯兰国(ISIS)成员汇聚模式的探索和发现^[3]。还有,如脸书运用社交数据进行的情绪传染实验和观点传播实验,都是很好的例子^[4]。非常遗憾的是,在英文文献中,从数据出发的知识发现研究很

[1]Cioffi-Revilla, Claudio, *Introduction to Computational Social Science*, Cham: Springer International Publishing, 2017.

[2]Ginsberg, J., Mohebbi, M., Patel, R., et al., "Detecting Influenza Epidemics Using Search Engine Query Data", *Nature*, 2009, 457, pp.1012-1014.

[3]Johnson N. F., et al., "New Online Ecology of Adversarial Aggregates: ISIS and Beyond", *Science*, 2014, 352(6292), pp.1459-1463.

[4]Kramer, Adam D. I., Jamie E. Guillory, et al., "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks", *Proceedings of the National Academy of Science*, 2014, 111 (24), pp.8788-8790. Bakshy, Eytan, Solomon Messing, and Lada Adamic, "Exposure to Ideologically Diverse News and Opinion on Facebook", *Science*, 2015, 348(6239), pp.1130-1132.

少有极好的文献发表在社会学主流期刊上,更多的是被当作自然科学研究的一部分发表在各类自然科学期刊上。在中文文献中,陈云松等运用微博和百度数据对代内文化反授的研究倒是极具计算社会学的品格^[1]。

归纳起来看,作为研究方法的计算社会学演进,一方面受到建模技术发展的影响,另一方面受到数据挖掘技术的影响,呈现出两条路径平行发展的格局。在这个格局中,一方关注自下而上的基于行动者的互动;另一方关注总体性关系关联。两种格局看似延续了传统社会学微观和宏观的方法路径,事实上却是为打通宏微观关系建设了两条相向而行的通道。从微观行动出发的 ABM 其实有能力刻画宏观的涌现或复杂性,从总体出发的关系发现也有机会收敛到具体场景的互动机制。只是当下的积累尚不厚,还有待更多对社会现象有兴趣的社会学家和数据科学家合作,共同面对数字技术带来的数字社会。

四、结论

经历了 40 年数字技术社会化应用的发展,数字社会已经进入本质性社会变革在各个领域发生和呈现的阶段。尽管 1993 年就有人提出了计算社会学概念,可是,20 世纪 90 年代以来,主流社会学家们显然逐步脱离了社会现实,而陷入对传统社会学的怀旧之中不能自拔,对日新月异的社会变迁表现出一种无力感,在研究方法上难以突破。

最近 10 年,随着数字数据越来越丰富,在数据科学蓬勃发展的挤压之下,社会学家终于有所醒悟,正试图正面对巨量数据不断积累、社会复杂性日趋显性的社会现实,发展出两条有可能互补的计算社会学研究路径。一条是自下而上、着眼于行动者主观能动性和社会行动动态性的 ABM 路径,主要用于机制发现和检验;另一条则是着眼于从巨量数据中发现社会模式或规律的数据挖掘路径,两者都运用了计算思维,即运用分布式算力、高维变量、巨量数据,以获取人类行为和社会现象规律的计算效率逻辑,主要用于模式或知识的发现与检验。值得注意的是,在两条路径上还是两拨人,一拨是以数理建模为基础的社会学家;另一拨是以计算建模为优势的社会学家。两拨人还没有合流,将来是否合流尚未可知。值得欣喜的是,不管怎样,社会学家总算是面对了时代的潮流。不仅如此,自然科学家也加入探索社会规律的潮流之中,这意味着,社会学家如果自己不做,便会有自然科学家替代之。

在拙稿停笔之际,《自然》杂志于 2021 年 7 月 7 日又出版了一期计算社会科学特刊,内容涉及计算社会科学的诸多内容,非常值得参阅。最后,需要说明的是,我们可以把计算社会学作为计算社会科学的一部分,可计算社会科学不能替代计算社会学,特别是不能替代其特定的研究领域和方法路径。当然,计算社会学与计算社会科学的同与异,不属于本文内容,冀另文探讨。

[责任编辑:方心清]

[1]陈云松、朱灿然、张亮亮:《代内文化反授》,《社会学研究》2017 年第 1 期。