

上篇

计算传播学：社会科学研究的新的取向

在这一篇章，本书将引导读者深入理解计算传播学的核心理念与方法，阐释这一新兴学科如何在社会科学领域发挥着越来越重要的作用。随着信息技术的飞速发展，尤其是大数据、人工智能等技术的普及，社会科学研究进入了一个全新的计算时代。传统社会科学方法，尤其是定性研究和基于小样本的定量研究，在面对海量数据时显得力不从心。计算传播学作为社会科学的重要分支，正致力于通过先进的计算方法与工具，帮助学者们从大数据中提取出更为精确和深入的传播规律。

第一章从“社会科学是什么”开始，梳理了社会科学的研究范围和核心目标。通过这一基础的理论框架，读者将了解到计算社会科学的诞生不仅是信息技术与统计学发展的必然结果，也是在传统社会科学方法面临挑战时的应运而生。本章将详细介绍计算社会科学的起源，如何通过计算机技术的引入，尤其是在数据收集、处理和分析中的应用，推动了社会科学研究的突破性进展。

第二章深入分析计算传播学的兴起背景与应用领域。随着社交媒体、互联网及大数据技术的普及，传播学研究不再局限于传统的问卷调查和定性研究，而是进入了一个数据驱动的新时代。本章将深入讨论计算传播学如何运用大数据、社交网络分析、情感分析等手段，解决传统传播学研究无法应对的新问题，如实时传播分析、跨平台信息流动和用户生成内容的影响力等。

第三章聚焦于计算传播学的研究设计，重点探讨了这一领域的研究方法与传统传播学量化研究方法的差异。计算传播学不仅是数据的简单统计，更重要的是如何将复杂的数据转化为具有实际意义的传播学洞察。本章将详细介绍如何通过统计学方法与编程工具相结合，设计出符合传播学问题的研究方案。从研究问题的提出，到数据收集、处理和分析，每一步都需要精准的工具和方法来支持。此外，本章

2 计算传播学：基础、方法与实践

还将探讨在计算传播学研究中常见的挑战,如数据的选择性偏差、算法透明度问题等,并给出解决方案,帮助读者更好地理解如何设计可靠的研究。

通过本篇的学习,读者能够理解计算传播学如何为社会科学研究提供新的视角和方法,掌握其核心概念和研究设计框架,为进一步的技术应用与实践探索打下坚实的理论基础。

CHAPTER 1

第一章

进入计算社会科学时代

第一节 社会科学是什么

一、自然科学、社会科学与人文科学

所谓“科学”(science),是一种系统性的知识体系,它积累、组织并可检验有关万物的解释和预测。科学不仅在于已经认识的真理,更在于探索真理的活动,即对自然、社会乃至人本身进行研究的整个过程。科学强调预测结果的具体性和可证伪性,这有别于以理性追问终极的哲学。科学也不等同于寻求绝对无误的真理,而是在现有基础上,摸索式地不断接近真理。判断科学与否的参照标准:可重复性与可证伪性。可重复性是指一个研究的过程与结论可以在不同的时间和空间里加以复制,这样可以证明规律的存在具有必然性,而不是偶然出现的。可证伪性是指从一个理论推导出来的结论,在逻辑上或原则上要有与一个或一组观察陈述发生冲突或抵触的可能,即这些结论必须容许在逻辑上有反例的存在。同时,科学也是一种社会职业和社会建制。作为知识体系的科学既是静态的,也是动态的——思想可以产生思想,知识在进化中可以被废弃、修正和更新。作为研究过程和社会建制的科学是人的一种重要社会活动。

科学知识是人类通过实践和理性对自然和社会现象进行探索和解释的有序知识体系。科学知识可以根据不同的依据进行分类,如研究对象、研究方法、应用目的等。其中,最常见的分类方法是按照研究对象的不同,将科学知识分为自然科学、社会科学和人文科学三大部类。自然科学主要研究自然界的物质、能量、生命等现象和规律;社会科学主要研究人类社会的结构、功能、变迁等现象和规律;人文科学主要研究人类的精神、文化、价值等现象和规律。这三大部类之间既有区别又有联

系,共同构成了人类对客观世界的认识。

一般而言,自然科学是实验科学或经验科学,也就是说它是基于经验的。自然科学的外延主要是物理科学和生命科学两大类,包括数学、物理、化学、地理学、生物学等。关于自然知识的历史是漫长的,但是真正意义上的自然科学的历史只有三四百年。

与自然科学不同,社会科学以社会作为研究对象,其目标在于运用科学的方法,认识并解释各种社会现象并尽可能找出它们之间的关联,发现社会发展规律。然而,社会科学与自然科学之间的差异并非天壤之别。在理论建构、研究方法等方面,社会科学都在一定程度向自然科学借鉴。事实上,社会科学中的某些学科已经运用各种数理工具,变得与自然科学毫无二致。因此,社会科学也是一种“科学”。

正如哈耶克所说,“社会科学研究的不是物与物的关系,而是人与物或人与人的关系。它研究人的行为,它的目的是解释许多人的行为所带来的无意的或未经设计的结果”。自18世纪启蒙运动兴起以来,社会科学逐渐发展壮大,现已涵盖经济学、政治学、军事学、法学、伦理学、社会学、心理学、教育学、管理学、新闻传播学、人类学、民族学等众多学科。19世纪末20世纪初,受欧洲社会变迁和自然科学发展的影响,社会科学成为独特的、公认的学科。它建立在对社会现象和规律的系统 and 理论认识的基础上。早期的社会科学代表人物有孔德、马蒂诺、边沁、斯宾塞、涂尔干、韦伯、马克思等,他们在不同领域提出了自己的理论和方法,为社会科学的发展奠定了基础。

二、社会科学研究的特殊性与两条进路

与自然科学不同,社会科学的对象是人,特别是由人组成的社会的运行和变迁规律。在结构-功能的关系讨论中,社会科学的许多研究结论并不如自然科学那般稳定,其社会规律的反馈机制也比自然科学复杂。因而,社会科学具有诸多特殊性。

第一,社会科学研究要考虑到社会历史现象中人的主体参与性。自然科学研究者能够在边界明确的环境中对特定对象展开实验研究,然而这一情况在社会科学研究中则不同。社会科学研究中的人既是社会产物,又是社会的创造者;既是社会主体,又是社会客体;人既受社会规律支配,又具有主体能动性。因此,认识、解释社会现象并发现其规律的过程中,无不体现着人的主体能动性。这就会影响社会科学研究中信息、资料的客观性,理论预测的准确性,也就使社会科学研究变得复杂。

第二,社会现象表现出个体性、偶然性、随机性。自然科学能够创造近乎相同的实验环境重复科学研究结果。这种研究结果的可重复性使得自然科学的知识更加“科学”而被公认。然而,社会科学研究很多时候难以摆脱具体的人文历史、社会思潮、文化语境、地域等方面的影响。同一个社会科学研究问题,在不同的国家、不同的时期可能会得到完全不同的结论。

第三,人类社会是一个极其宏大的系统,其中包含的要素参量、变量甚多。人类社会是由人与人之间相互作用形成的复杂组织,它具有动态性、多样性和可变性。社会现象是指人类社会发生的各种事情和活动,它们反映了人们在特定时空条件下的思想、行为和关系。自然现象则是指自然界中发生的各种事物和过程,它们遵循着一定的物理、化学或

生物学规律。研究社会现象不能像研究自然现象那样,用一般的类比、归纳的办法来处理。例如,我们可以通过观察和实验来归纳出水在不同温度下的状态变化,但我们不能用同样的方法来归纳出人们在不同情境下的心理变化。因为社会中的个体差异是繁杂而悬殊的,这种个体差异不仅取决于生物学上的原因,更取决于经济、政治、文化和心理等复杂的社会因素。因此,研究社会现象需要运用更为灵活、综合和创新的方法。

第四,相较于自然科学,社会科学具有社会意识形态性。一般来说,社会科学中的部分学科属于意识形态和上层建筑范畴,它们直接或间接地反映社会经济基础和政治制度的要求和需要。因此,这类社会科学研究更可能受研究者的阶级立场、利益倾向、价值取向以及所处客观条件等方面的制约。

造成上述这些区别的根本原因在于,自然科学研究的对象是无意识的、无目的活动的客观过程或物质客体,而社会科学研究的则是现实的、活动着的个人,这就决定了社会科学研究的复杂性、困难性和难预测性。

近年来,社会科学也发展出了许多新的特点,如社会科学与自然科学交叉融合。各社会科学学科之间也相互渗透,这使得传统思辨哲学对社会科学的影响逐步弱化,社会科学开始注重探讨社会现实问题。社会科学的研究者们开始将如社会网络分析、自然语言处理、机器学习、基于个体的模拟建模(agent-based modeling)等前沿研究方法运用到社会科学领域的研究中。这使得社会科学不局限于“过去的研究”,同时也开始展开面向未来的研究,进行社会预测和积极干预。

面对复杂的社会问题,社会科学家们发展出了两条研究进路,分别是实证主义和人文主义。通常意义上,现代社会科学的研究历史可以追溯到17世纪。以奥古斯特·孔德(Auguste Comte)为代表的法国社会科学家提出了社会系统的自然科学,完善了数学及统计的基础和方法,以此改进了传统的历史方法及早期的哲学方法。自那时候起,社会科学的发展产生了大量理解人类及社会行为的知识,也随之诞生了所谓五大社会科学学科,即人类学、经济学、政治学、心理学和社会学。

孔德提出了实证主义(Positivism),论证了自然科学方法应用于社会研究的合理性。实证主义起源于培根的经验主义和牛顿-伽利略的自然科学方法,是社会科学研究中常见的一种研究范式。孔德区分了实证主义与英国的经验主义的区别,他认为,经验主义虽然也强调观察和验证,但它只是简单地堆砌零碎的事实,实证主义的不同之处在于它还要运用思辨的逻辑揭示永恒的普遍规律,做出预测。

实证主义认为,社会现象有着其本身的内在规律和发展顺序,这些普遍规律可以重复发生;人是理性的个体,受到外部世界的刺激而做出相应的活动,同一个刺激在不同人的身上会产生相同的结果,因此,通过观察人的行为,就可以理解他们、推断他们的行为;外部刺激是原因,人的行为是结果,研究正是要寻找这种普遍性因果关系。因此,实证主义研究范式的社会科学家们常采用量化研究方法,通过数理统计、建模、实验等方法开展研究。

然而,考虑到上述社会科学研究的特殊性,要想实现绝对的、完全的客观是不可能的。如果将人的思维情感进行高度抽象,那便与石头无异。但是这并不意味着社会科学研究

不需要一套完整的研究标准,而是需要社会科学家在研究过程中不断反思。

在社会科学研究中,与实证主义相对应的研究范式是人文主义,也被称为解释性的研究范式。人文主义拒绝实证主义将人高度抽象的工具论取向,它肯定人性和人的价值。人文主义认为人的活动是自愿的,需要在具体的情境中理解人类的行为,而研究的目的在于理解行为背后的意义,强调得到一个对人类行为和社会现象的整体性认识,而不是通过所谓科学的方法预测或干预控制人的行为。这种研究范式认为知识是被社会建构的,同时也是历史性的。这就意味着并不存在一个永恒不变的社会定律,随着时间流逝,社会理论可能会过时。社会出现新的现象,已有的理论无法对此进行有说服力的解释时,就需要社会科学家提出新的理论。同时,人文主义承认研究者本身在研究过程中是有价值偏见的,也认为所谓真实也不是唯一的,而是人们相互建构出来的。对研究结果的检验标准,是看研究对象对此的理解以及帮助其他人理解研究对象。

在社会科学中,特别是传播学,同时存在着多种研究范式。在社会科学的研究中遵循何种范式是一个实践问题,与研究者、研究问题、研究条件等方面都相关。没有哪一个范式能够回答所有的研究问题;相反,不同的范式之间相互竞争,才能共同促进学科的繁荣发展。

第二节 计算社会科学的诞生

“范式”这一概念最初由美国著名科学哲学家托马斯·库恩(Thomas Kuhn)于1962年在《科学革命的结构》中提出,指的是常规科学所赖以运作的理论基础和实践规范。范式是从事某一科学的科学家共同体所共同遵从的信念、基本理论、价值、技术等一整套系统,包括判断研究问题是否重要的标准、判断研究与社会关系的标准等。范式的意义在于维持研究的延续性,能够对科学家共同体起规范作用,协调他们对世界的看法及其行为方式。上一节中提到的实证主义和人文主义,即可理解为社会科学中的两种研究范式。

在范式的基础上,库恩认为,任何一门科学知识的发展都是一个演化与革命、积累与创新、连续与间断交替发生的历史过程。一般来说,科学发展存在四个阶段,分别是前学科(没有范式)——常规科学(建立范式)——科学革命(范式动摇)——新常规科学(建立新范式)。科学家在没有共同范式的前学科时期,意见不一。不同范式之间的竞争和选择,最终导致一个范式获得科学共同体的认可,进入常规科学时期。在这个时期,科学家遵循范式进行释疑活动,推动科学进步。然而,释疑活动也会带来新问题和新事物,对原有范式造成挑战,引发新范式的科学革命。革命使新范式取代旧范式,形成新的科学共同体。新范式并不意味着更接近真理,只是对现实有更好的解释能力,开启新的常规科学。

千百年来,科学研究经历了范式演化过程并产生了四大范式,分别为经验科学、理论科学、计算科学和数据密集型科学(也被称为大数据科学)。

早期科学以记录和描述自然现象为主,可称为“经验科学”,也可称为“实验科学”,即第一范式,其典型案例如“钻木取火”。经验科学偏重于经验事实的描述和明确具体的实

用性的科学,一般较少抽象的理论概括,在研究方法上以归纳为主。一般科学的早期阶段属经验科学,化学尤甚。事实上,只要自然科学和社会历史科学是从感觉经验出发的,都是以经验材料为其研究对象,都具有经验科学的性质。

文艺复兴以来,科学家们开始利用模型归纳总结过去记录的现象,发展出“理论科学”即第二范式,其典型案例如牛顿三定律、相对论等。科学第二范式是指以建模和归纳为基础的理论学科和分析范式,又称为理论范式。它是在自然现象的基础上进行抽象简化、主要通过构建数学模型来进行科学研究的一种范式。例如,相对论、博弈论等都是属于理论科学的典型代表。理论科学强调高度普遍的理论认识,而不是直接实用的科学。

过去数十年间,计算机的出现催生了科学第三范式,即“计算科学”。在第三范式中,计算机仿真模拟取代实验,运用数据模型构建、定量分析方法以及利用计算机来分析和解决科学问题,反映了科学从理论主义向计算主义的转变。计算科学可以对复杂现象进行模拟仿真,重建和理解已知事件(如地震、海啸和其他自然灾害);预测未来或未被观测到的情况(如天气、亚原子粒子的行为),其典型案例如模拟核试验。此外,第三范式更加倾向于寻找相关关系,而不是因果关系,更加关注“是什么”而不是“为什么”。

大数据时代,科学发展的新趋势是利用计算机对海量数据进行分析总结,发现规律,形成“数据密集型科学”或称第四范式。该研究范式不再需要模型和假设,而是通过分析海量数据进而获取知识。第三范式与第四范式的区别在于:前者以可能的理论为先导,进而搜集数据,然后通过计算机仿真验证理论;后者则是在占有大量数据基础上,通过计算得出未知的理论。

著名学者、1998年图灵奖得主、关系数据库的鼻祖吉姆·格雷(Jim Grey)关于第四范式的论说给世人留下了深刻的启发,大数据、人工智能等技术革新都印证了他的预言。出版于2009年的《第四范式》(*The Fourth Paradigm*)^①一书系统介绍了基于海量数据的科研活动、过程、方法和基础设施,揭示了数据密集型科学与实验科学、理论科学、计算科学这三种科研范式相辅相成的关系。数据密集型科学强调要更多地利用数据,在科研领域发展并推广数据相关的技术;要促进数据和科研成果的分享和更广泛的使用,因此,要促进研究者之间的新型合作,这将是一种大规模的、跨学科的、高频率的合作。

第三节 计算社会科学的发展

从实证主义出发,传统的社会科学研究多采用问卷调查、线下实验等研究方法,通过逻辑严密的理论推导和数理统计来回应研究问题。然而,传统的社会科学研究在经验数据的获取、数据分析与因果验证方面仍存在一定局限性。

随着科学技术的不断发展,特别是移动互联网的崛起,这一局面正在慢慢发生以下改变。第一,如今我们生活在一个网络时代。我们在公共场所的活动可能会被监视器捕捉;我们的医疗记录会被存储为数字文档;我们可以发布任何人都可以访问的微博,或者通

^① Tansley S, Tolle K M. The fourth paradigm: data-intensive scientific discovery[M]. Redmond, WA: Microsoft research, 2009.

过微信朋友圈记录自己的日常生活。人们的日常行为都可能被数字化地记录下来,成为数字踪迹(digital trace)。2009年,拉泽(Lazer)等学者^①在《科学》(*Science*)期刊上发表《计算社会科学》(“Computational Social Science”)一文。该论文被计算社会科学研究者视为“宣言”,文章指出,曾经因为收集和分析大量数据能力的发展,生物学和物理学等领域得到飞速发展;如今利用计算方法和技术来分析及理解人类传播过程及现象,将可能彻底改变我们对社会生活、组织活动等各个领域的理解。

那么何为计算社会科学呢?根据美国计算社会科学的定义,计算社会科学(computational social science)指的是在社会科学中采用计算机运算方法的学术分支,运算用以建立模型,模拟、分析社会现象,包括计算经济学、计算社会学、历史动力学、自动媒体分析等。计算社会学科是一门严肃科学。科学解释意味着提供一种因果机制来解释观察到的现象,如事件特征和规律。因果机制可以是确定性的,也可以是概率的。作为一种实证科学,计算社会科学家需要对社会进行观察和测量,其研究任务通常是用数学描述社会现象如何发生,揭示其中的社会定律。

如今计算社会科学在传统社会科学的各学科中都产生了广泛的影响,但学者们大多不认为计算社会科学与传统社会科学有着巨大不同;相反,它们关系紧密。很多时候,它们只是对同一个问题,提出了不同的解决方案。计算社会科学研究者需要从传统社会科学的研究中汲取理论养分,提出有洞察力的研究问题;而传统社会科学的研究问题经过进一步的验证之后,能够为理论提供更充分的论据支持,使其焕发出更强大的现实解释力。

事实上,在拉泽等发表“宣言”一文以前,研究者就基于计算的方法,如数学模型、复杂网络分析等工具,发现了诸多经典社会科学理论,如六度分隔理论(six degrees of separation)、弱连接(weak ties)、结构洞(structural holes)、无标度网络(scale-free network)。这类通过数据建模以观察社会的研究也被称为社会物理学研究。然而,囿于数据、算法、算力等方面能力的缺乏,相关研究并没有得到足够的重视。2016年,迈克尔·阿尔瓦雷斯(R. Michael Alvarez)编著的《计算社会科学:发现与预测》(*Computational Social Science*)^②全面回顾和概括了计算社会科学的发展状况,发现计算社会科学拥有丰富的工具箱,并广泛应用于社交媒体、抗争运动、议会表决、新型政党组建、政府治理以及社会营销学等领域。

大数据、人工智能、云计算等技术的发展,为社会科学研究提供了新的数据源、方法和工具。学者能够以全新方式和全新视角重构和解读社会学研究的经典概念。面对社会不平等、传染病传播、政治行为、社会运动和社会网络等重要的社会现象和问题的研究,计算社会科学为解决实际问题提供了新的视角和策略。

总体而言,计算社会科学的发展大致受到四个方面的作用:第一,传统议题与新兴议题的结合。计算社会科学既关注新数据、新技术对人类世界带来的新问题、新挑战,如网络舆情、群体事件、社会物理等;又致力于以全新方式和全新视角重构和解读社会学研究

^① Lazer D, Pentland A, Adamic L, et al. Computational social science[J]. *Science*, 2009, 323(5915): 721-723.

^② Alvarez R M. *Computational social science*[M]. Cambridge University Press, 2016.

的经典概念,如阶层 / 阶级、社会流动、社会观念等。第二,主体性与群体性的平衡。计算社会科学的灵活性和适应性有助于从点到面地对人与群体的互动展开研究,既重视个人行为属性分析的基础性地位,又强调社会关系、社会网络及群体特征分析的宏观架构。第三,外部条件性与内部动力性的互动。计算社会科学不仅强调社会生活中信息规律和制度演变等外部条件的探讨,如由信息内容分布的解析到话题发现和言论传播;也关注这些外部条件与人类个体心灵之间相互作用、相互影响的关系,如个体情感、认知、态度等。第四,独立性与交互性的协调。计算社会科学不仅通过对范式和话语体系的重构来为以往社会学研究中的模糊概念“划界”,也强调不同概念及不同概念所代表的社会现实之间的交互性,如个体的倾向性、可信度和影响力等属性对外界条件的响应模式、传导机理和交互适应性,以及关系网络演变。

博雅斯卡亚(Buyalskaya)等学者^①在《美国国家科学院学报》(*Proceedings of the National Academy of Sciences, PNAS*)上发表了论文《社会科学的黄金时代》(“The Golden Age of Social Science”)。作者发现新型的数据及其分析方法、跨学科合作正在出现爆炸性增长。计算社会科学将计算机与信息通信技术运用于社会科学研究,致力于克服传统实证研究方法的局限,为社会科学研究开辟新的路径,正进入一个黄金时代。计算传播学正是计算社会科学中的重要一环。

^① Buyalskaya A, Gallo M, Camerer C F. The golden age of social science[J]. *Proceedings of the National Academy of Sciences*, 2021, 118(5): e2002923118.