

# 数字化与文科范式革命

邱泽奇

(北京大学社会学系, 北京 100871)

**摘要:**数字化是人类经历的第三轮技术革命,正在席卷社会每个角落,也是文科正在进行的第三轮范式革命的推动因素。文科范式指从观念到实践、从教学到科研、从学术到知识的逻辑一致性体系;既具有库恩范式的基底,更指文科的资源/素材、方法/技术、理论/视角的关联一致性体系。文科的第一轮范式革命是学术革命,第二轮是计量革命,第三轮则是正在进行的智能革命。学术界和企业界正在实践的数智文科则是从素材像素化到理论智能化的探索式知识创造与积累。国家倡导和推动的新文科建设是文科范式革命在中国的组织化实践。如果我们把正在展开的文科范式革命划分为四个阶段(素材像素化、像素数据化、数据理论化、理论智能化),那么,当下的努力其实还处在第一、二阶段,而第二、三、四阶段则可能是交互并进的。

**关键词:**数字化;文科范式;知识生产

数字技术革命是迄今为止人类发生的第三轮技术革命,正在改变社会的生产与生活,乃至一切。农业技术革命以动物养殖和植物栽培为标志。在中国,栽培水稻的历史可上溯至14 000—18 000年前,《史记·夏本纪》有“禹令益予众庶稻可种卑湿”,在湖南省玉蟾岩遗址发现的水稻种子约在14 000年前。<sup>①</sup>农业技术革命为人类带来食物,让人类延绵至今。工业技术革命则以机械动力进入生产实用为标志。在欧洲,以纽克曼(Thomas Newcomen)1712年发明蒸汽机为标志,拉开了动力技术革命的序幕,历经化石燃料和电力动力的变革,彻底改变了社会的生产和生活方式。<sup>②</sup>工业技术革命为人类带来充裕的物质,让人类可以衣食无忧。数字技术革命以可编程电子计算机的诞生为标志,以互联网的大众化应用为转折点,<sup>③</sup>是人类正在经历的、从一开始便影响广泛且深远的未有之变。

作者简介:邱泽奇,北京大学中国社会与发展研究中心主任、社会学系教授。

① 在玉蟾岩遗址和河姆渡遗址考古发现水稻种子的事实已有广泛报道,故不再注明出处。

② 讨论工业技术革命给人类影响的文献亦汗牛充栋。初期部分可参见保尔·芒图,《十八世纪产业革命:英国近代大工业初期的概况》,杨人梗、陈希秦、吴绪,译。北京:商务印书馆,1983。

③ 世界上第一台可编程电子计算机恩尼克(ENICA)于1945年12月10日在美国宾夕法尼亚大学投入工作。世界上第一个面向公众的网络浏览器于1993年由美国国家超级计算机应用中心(NCSA)发布,即马赛克(Mosaic)浏览器。

数字化带来的影响如此之深刻,以至于数字化已经成为探讨诸多问题时的必要前提。非数字化是一种场景,人们会以自然状态和机械状态为底色进行探讨;数字化则是另一种完全不同的场景,人们会以数字连接泛在为底色进行探讨。

科学研究是人类生产的重要组成部分,同样受到了数字化的影响。卷积神经网络技术在蛋白质结构研究领域的应用(AlphaFold 系列)彻底改变了这个领域的知识创造。<sup>①</sup>非数字化的与数字化的科学研究的差异可以用一项思想实验作为示例,即姚檀栋的第二次青藏高原科学考察进展报告。姚檀栋认为,第二次青藏高原科学考察“提出了青藏高原生态保护立法科学建议,服务高原生态文明高地建设;阐明了气变影响下亚洲水塔失衡特征和影响,服务国家水资源与水安全战略;揭示了气变影响下高原碳汇功能和变化特征,服务应对气变和实现双碳目标”等,共获得了十大重要进展。<sup>②</sup>姚檀栋对科学进展的报告呈现了非数字化场景的科学研究样态,即以学科化的成果代表科学研究的重大进展,每一项进展都限定在某个学科如生态学、水科学、气候学等边界范围内。如果把第二次青藏高原科学考察的成果汇集放在数字化场景里,十大重要进展或许会共同指向一个具有智能化理论产出的综合进展,如青藏高原与人类发展有怎样的关系。众多已经发生的和正在发生的事实表明,数字化对科学研究的影响是革命性的,<sup>③</sup>以至于不得不重构人类的知识创造和积累范式,而这个范式重构不是指转向人们已经广泛知晓的第四范式。<sup>[1]</sup>

## 一、文科的知识生产范式

说到“范式”,我们不能回避库恩(Thomas S. Kuhn)。库恩在他影响深远的著作中,对范式进行了定义。<sup>[2]</sup>

范式,通常是指那些公认的科学成就,它们在一段时间里为实践共同体提供典型的问题和解答。

这些著作之所以能起到这样的作用,就在于它们共同具有两个基本的特征:它们的成就空前地吸引一批坚定的拥护者,使他们脱离科学活动的其他竞争模式。同时,这些成就又足以无限制地为重新组成的一批实践者留下有待解决的种种问题。

凡共有这两个特征的成就,我此后便称之为“范式”。

在总共 13 章的篇幅中,库恩还呈现了科学范式的迭代机制(见图 1)。综合图 1 和库恩对

---

<sup>①</sup> 有关蛋白质结构预测系统的相关事实,参见 Jumper J, Evans R, Pritzel A, et al. Highly Accurate Protein Structure Prediction With AlphaFold. *Nature*, 2021, 596(7873): 583-589; Senior A W, Evans R, Jumper J, et al. Improved Protein Structure Prediction Using Potentials From Deep Learning. *Nature*, 2020, 577(7792): 706-710.

<sup>②</sup> 在 2023 年 4 月 18 日中国科学院学部举办的科技伦理研讨会上,姚檀栋院士报告了第二次青藏高原科学考察的重要进展。

<sup>③</sup> 参见如 Borgman C L. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press, 2007. 不过,这本书的讨论依然局限在有限的范围。

范式的定义,我们可以提出一个非常直接的疑问:知识创造与积累到底是对真理的追寻,还是对市场的占有。如果知识创造与积累不是追寻真理,那么,范式迭代便不过是产品更新的代名词而已,科学研究对人类的意义和价值便与商业产品的生产与更新无异,毫无神圣性可言。当然,这不是本文的主题,可以按下不表。一个更加直接的引申追问是,科学知识到底是理论的,还是经验的。在库恩那里,知识显然是理论的,他的底层假设承接了柏拉图(Plato)对先验(world of being)的断言。与库恩持相同立场的还有费耶阿本德(Paul Feyerabend)。他们从知识的先验性出发,认为人类的知识创新与积累不过是对先验知识的发现而已,在某种意义上,后者的观点甚至更为相对主义。<sup>[3]</sup>而在库恩之前的波普尔(Karl Popper)那里,知识则是经验的,<sup>[4]</sup>其底层假设承接了柏拉图对后验(world of becoming)的言说。与波普尔持相同立场的还有他的学生和批判者拉卡托斯(Imre Lakatos)。<sup>[5]</sup>他们从知识的经验性出发,认为人类的知识创新与积累是线性的,是从低阶向高阶、从简单到复杂的演进。

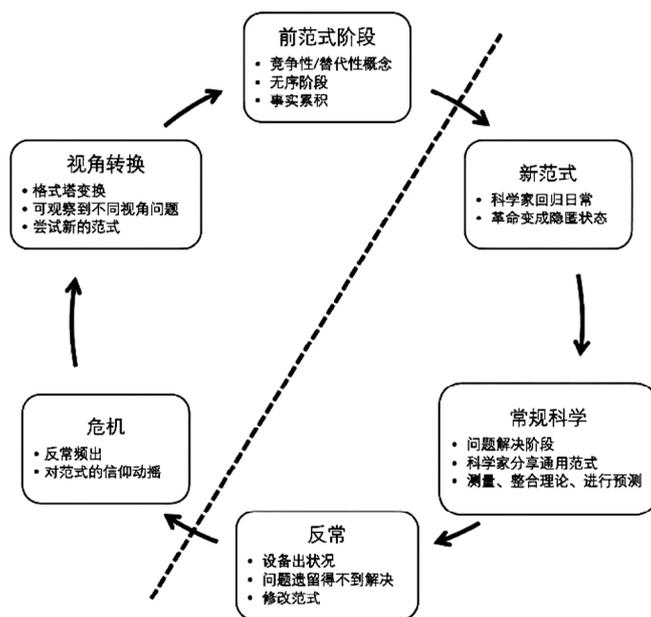


图1 库恩的范式迭代

既有对人类知识创造与积累的探讨,依然留下了一系列的问题。一方面是一些宏大的议题,如人类知识是一个整体,还是一个组合,人类知识到底是进化的,还是革命的,人类知识是科学家的筛选,还是社会的选择,等等。另一方面则是一些具体的讨论,如库恩的范式到底是人类知识创造与积累的范式,还是只是一个学科知识发展的范式,如果是一个学科知识创造与积累的范式,那么科学是人类知识创造与积累的唯一选择吗?

我们并不试图回答那些宏大问题,也不准备纠结人类知识的来源,而是假定科学是人类创造与积累知识的有效路径,并且把库恩的范式放到库恩写作的时代背景下,认为库恩的范式在严谨的意义上其实只是一个学科的知识创造与积累范式。然而,从库恩的范式中,我们可以归纳出范式变革的要素,将其用于理解人类知识创造与积累的范式。不过,本文只打算探讨文科的范式革命。

人类的知识创造与积累离不开三个基本要素:素材、方法、视角。库恩对范式的探讨可以被梳理为一个这样的框架:在给定的学科,给定研究素材,给定或改变研究方法或技术所带来的知识变革。前面的图1可以被理解为这个框架的逻辑呈现。我们的问题是,如果对知识的创造和积累不是局限于某个学科,而是整个文科呢。

人们通常理解的文科包含着人文学科和社会科学两个大的类属。当然,对人文是否应该或可以被称为科学,有着不同的争议性观点,且争议一直存在。<sup>①</sup>我们无意介入争论,只是从知识创造与积累的范式入手探讨范式革命。我们理解的文科是,人文学科聚焦对人类精神的挖掘与呈现,社会科学则着力于社会规律的发现与呈现。这里讨论的文科范式,既有库恩范式的基底,更涵盖科学活动与社会的关系。如果我们还接受库恩所述范式变革隐含的市场竞争逻辑,则文科范式除了库恩的学科范式之外,还有一个面向社会的知识生产与关乎合法性的关系范式。换句话说,知识创造与积累是学术界的事,而知识应用则是社会的事。如果只有知识创造与积累,没有知识应用,知识生产便失去了社会意义,也失去了社会合法性。无论是农业革命、工业革命,还是数字革命,带来革命的不只有知识的创造与积累,更加重要的还有对知识的社会应用。无论是人文学科还是社会科学,我们认为的文科范式还指从观念到实践、从教学到科研、从学术到知识的生产与应用的逻辑一致性体系。不过,这也不是本文关注的议题。我们关注的是因知识创造与积累的三要素(见图2)变革带来的文科范式革命。我们认为,在文科领域,知识生产是三个要素的关联一致性体系。任意要素的变动都会影响其他要素,也会影响关联一致性体系的结构稳定性,进而引发其他要素的变动。当三个要素的变动产生相变时,便会发生范式革命。当我们用这个框架分析文科范式的发展时,可以发现,人文学科与社会科学知识的创造与积累经历了三轮范式革命,当下我们正处在第三轮范式革命之中。

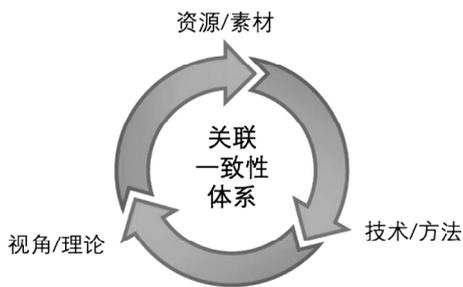


图2 文科范式要素

## 二、文科范式的三轮革命

文科的第一轮范式革命可以被称为学术革命。赫拉利(Yuval Noah Harari)将人类理性的出现称为认知革命,<sup>[6]</sup>其实,认知革命远比赫拉利讨论的要复杂。<sup>[7]</sup>我们说的学术革命是人类

<sup>①</sup> 库恩就是这个争议的代表人物之一。此外,还可以参见,Høyrup, Jens. *Human Sciences: Reappraising the Humanities Through History and Philosophy*. New York: SUNY Press, 2000.

运用理性认识自然和社会,创造和积累知识的范式革命。

在文科的第一个范式阶段,给定自然和社会是人类渴望认识的对象,被用于认识的素材除了实物,还有文本和图像;认识素材的方法是运用人类理性进行思辨和逻辑推演,其中,思想实验是最直接的也是最有效的技术手段,最终形成了对自然与社会的可积累的知识,即学术(academy)。

之所以称其为革命,是因为学术的出现将人类对自然与社会的认识划分为两个截然不同的类型。在学术革命之前,人类对自然与社会的认识没有文本图像的记录,只有语言或语音。这意味着人类对自然与社会的认识只能依靠人与人之间口耳相传。在人类的认识中,即使有知识,传播的范围也非常有限;即使有知识创造,也没有可靠的知识积累。文本图像的出现为知识创造和积累提供了可以留存与传递的载体,思辨或逻辑则让一部分人对自然与社会的认识从纯粹的超然信仰中摆脱出来,变成了有事实依据的认识过程,形成了可以被验证的知识,进而也让人类对自然与社会的认识从神谕中走出来,变成了学术活动的成果。

纵观人类历史,每一个地区或民族都有自己的学术革命。在中国,我们有诸子百家;在西方,有古希腊三圣。如果只是追溯以对象分科的现代学术源头,我们可以亚里士多德为例。亚里士多德是古希腊三圣中的最后一位,也是集大成者,人们通常称他为哲学家。殊不知他更是百科全书式的科学家。他是最早研究形式逻辑的人,系统研究物质运动的人,系统研究生物学的人,最早进行光学实验的人,最早进行太阳与地球关系探讨的人,最早进行地质探讨的人……现代科学的几乎每一个学科都可以从他那里找到源头。

在文科的第一个范式阶段,<sup>①</sup>面向社会的人类知识创造与积累是每一位学者独立的、个性化的知识生产过程。在芸芸众生中,一部分个体运用思辨和逻辑探索自然的规律和人类的精神,推动了从自然与社会的神圣性和神秘性向人类理性的演化。其中,无论因何种机缘而被认为承接了神圣的人,都获得了崇高的甚至超过君王的社会地位,成为那个时代的圣人。因此,那个时代的知识创造与积累,本质上是学者个性化地对自然与社会的属性进行的非计量、断句式、教导式、说理式的刻画与解释活动。人类获得的是精英化的知识。

在那个阶段,不仅人文学科与社会科学没有分离,甚至文科与自然科学也没有分离。具体地说,精英化的知识是对自然与社会的整体性认识。直到科学革命,才有自然科学从整体性学术的分离,形成了以自然的数据为素材的学术和以社会的现象为素材的学术。这一历史性分野的标志是1543年两部著作的面世。那一年,哥白尼(Mikołaj Kopernik)出版《天体运行论》(*De revolutionibus orbium coelestium*),维萨里(Andreas Vesalius)出版《人体构造》(*De humani corporis fabrica*)。两部著作只是对历史性分野的宣示,真正引发分野的是支撑两部著作的技术或方法与素材互动的变革。我们知道,无论是文科还是其他学科,人类进行知识创造与积累的对象都是自然与社会。面对同样的对象,之所以形成了分野,是因为哥白尼运用了计量观察技术获得了有关天体运行的新的数值型素材,且运用了计算技术处理素材;同样,维萨里运用

<sup>①</sup> 准确地说,是人类知识创造与积累的初始阶段,那时尚未分科,自然也就没有文科和其他的学科分野。如后所述,文科是在后来发展中才从对自然与社会的整体性认识中分离出来的。

了不曾有过的结构解剖技术获得人体构成素材,运用了计量技术分析素材。素材获取技术的变革,带来了素材之于理解和认识的意义和价值的变革,引发了分析技术的变革;分析技术的变革又进一步推动了素材获取技术的变革;两者之间的互动推动了对自然与社会的新的认识,推动了理论和视角的变革;最终推动的是人类知识创造与积累的变革。

在第一个文科范式后来的发展中,一个里程碑式的潜在变革是统计技术的引入和应用。在面向社会的整体性知识创造与积累中,统计思维和统计学把社会科学从整体性学术中分离出来。1663年,格朗特(John Graunt)发表《对死亡率清单的自然和政治观察》,标志着对社会现象的观察获得了数值型素材,而“死亡率”这一统计量的出现意味着将统计思维引入处理数值型素材的过程中。之后,伯努利(Jacob Bernoulli)于17世纪末期孕育概率论,高斯(Carl F. Gauss)于1795年运用最小二乘法,最终让统计学成为知识创造与积累的新工具。高尔顿(Francis Galton)和皮尔逊(Karl Pearson)创立的生物统计学进一步标志着现代统计学在知识创造与积累中进入实用阶段,也标志着素材技术和分析技术对第一个文科范式产生了实质性的侵蚀。

社会学正是在第一个文科范式的被侵蚀中分离出来的社会科学。提到社会学,人们马上想到的一定是孔德(Auguste F. X. Comte)。不过,如果从史料论来看,孔德只是提出了社会学概念而已。公允地说,他的社会学还是哲学里的一个倡导。让社会学从整体性学术中分离出来的是涂尔干(Émile Durkheim),涂尔干在《自杀论》(*Le suicide étude de sociologie*)中对自杀现象的数值型观察和对观察记录的统计学分析让他的自杀研究彻底区别于对自杀现象的神学或日常解释,带来了自杀率的科学分析。让社会学从整体性学术中分离出来的还有马克思(Karl H. Marx)。马克思在《资本论》(*Das Kapital: Kritik der politischen Oekonomie*)里不仅运用了统计学,还运用了数学建模。素材技术和分析技术的变革共同推动了理论的变革,涂尔干的功能论和马克思的冲突论正是运用新方法分析新素材后提出来的,也因此奠定了让社会学从整体性学术分离的学科范式。文科范式变革的新三角(见表1)标志着文科第二轮范式(计量)革命的开始。

表1 第二个文科范式

要素	人文学科	社会科学
素材	文本、图像、实物 言语、音乐、舞蹈、仪式、行动	文本、图像、实物 言语、音乐、舞蹈、仪式、行动 统计数据、调查数据
方法	比较分析、想象延展	比较分析、想象延展 思想实验、实证分析
理论	价值、意义阐释	价值、意义阐释 规律检验

计量革命是以计量技术为内核的第二个文科范式。计量革命引发了从对象细分到资源/素材、技术/方法、视角/理论的系统性变革,推动了整体性学术向分科性知识创造与积累的发展,其中包括现代社会科学。这里,我们尝试用一个比喻来刻画从整体性学术到现代文科的变革。我们可以想象图3的左边是整体性学术,右边则是现代文科(见图3)。

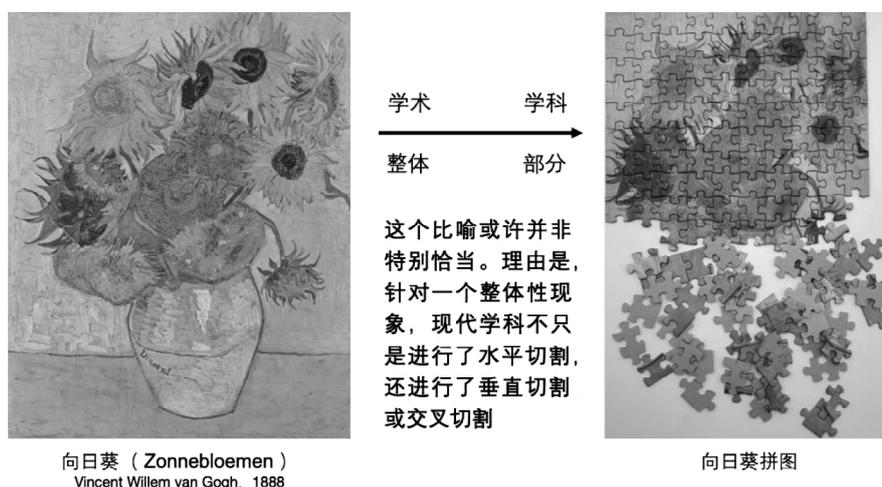


图3 文科的分科化

计量革命带来的是，把整体性对象切分为分科性对象，让整体性学术转变为学科分异的对象拼图。就像解剖学把有机生物肢解为肢体和器官一样，每个肢体或器官在新的学科中变成了完整的对象，却不一定是有机整体的部分，也因此有了面向社会的不断分化的学科。

在文科从整体性学术向分科性学科发展的进程中，社会学其实不是第一个分离出来的学科。第一个是经济学。斯密(Adam Smith)的《国富论》(*An Inquiry into the Nature and Causes of the Wealth of Nations*)标志着古典经济学的开端。社会学也不是最后一个分离出来的学科。在社会学之后，还有政治学。尽管有些人认为马基雅维利(Niccolò Machiavelli)的《君主论》(*Il Principe*)标志着现代政治学的开端，但在我们看来，那还是整体性学术的遗存，标志着现代政治学肇始的还是米歇尔斯(Robert Michels)对政党的分析。<sup>[8]</sup>

计量革命并不限于社会科学。在后来的发展中，运用计量方法获得新素材、分析新素材的范式也对人文学科进行了侵蚀。经典的人文学科如历史学也在运用计量方法，如考察地球上的人类联系，探索现代社会的起源和发展，<sup>[9]</sup>还有专门的期刊，如《计量史学：计量历史与文化演化杂志》(*Cliometrics: The Journal of Quantitative History and Cultural Evolution*)；科技考古也不断将计量方法引入对遗址和文物的分析，形成了诸如考古化学、考古地质学、考古材料与技术等新学科的发展。<sup>[10]</sup>即使是哲学，也有运用计量方法的新学科，如分析哲学。自弗雷格(Friedrich L. G. Frege)提出概念演算(Conceptual Notation)以来，分析哲学便形成了以数理逻辑为基础的发展路径，维特根斯坦(Wittgenstein)的《逻辑哲学论》(*Tractatus Logico-Philosophicus*)进一步确认了这一路径。

计量革命是一个逐步将计量技术与方法向知识的创造与积累渗透并形成范式优势的过程，尽管整体性学术依然存在，但其优势地位在逐渐消失，进一步形成了整体性学术与分科性学科之间的分野。与此同时，通过前文提到的知识生产与知识应用的关系，科学家的知识创造与积累经由制度化的教育体系和产业生产体系，变成工作岗位、生产和生活用品，进入大众生活，形成科学普及，让第一个范式的精英化知识变成了与社会密切相关的世俗化知识。

文科分科化对社会认识的效果与自然科学分科化对自然认识的效果不同。如果说自然现

象具有结构一致性,正如物理学对物质属性的探索可以从天体逐渐细粒化到量子,<sup>①</sup>社会现象却具有有机性,社会始终是一个互动的整体,是一个人类生态,而学科化的人文学科与社会科学将一个有机整体和生态切分为一幅幅拼图,无法形成对社会整体的有效认识。这便是科学革命以来文科发展的内生困境。在过去的时间内,文科学者们进行了各种努力,引发各种争议讨论,却始终没有找到解决内生性困境的有效路径。

数字化带来的变革对文科而言既是挑战也是跳出分科化困境的机会。学科史的事实告诉我们,计量革命带来的素材变革是部分的,是人力所及的计量化,非计量化素材依然支撑着文科的知识创造与积累。数字化则意味着一切素材的计量化,进而意味着把人文和社会的一切属性转化为数值数据,为将分科化的知识整合为对人文与社会整体属性的知识提供了机会,也意味着文科范式第三轮革命的到来(见表2)。

表2 第三个文科范式

要素	人文学科	社会科学
素材	文本、图像、实物 言语、音乐、舞蹈、仪式、行动 统计数据、调查数据 生成数据、一切皆可数据化	文本、图像、实物 言语、音乐、舞蹈、仪式、行动 统计数据、调查数据 生成数据、一切皆可数据化
方法	比较分析、想象延展 思想实验、实证分析 拟合与预测	比较分析、想象延展 思想实验、实证分析 拟合与预测
理论	价值、意义阐释 规律检验 规律发现	价值、意义阐释 规律检验 规律发现

比较表1和表2,如果说表1的人文学科继承了第一个范式整体性学术的传统,而社会科学是对整体性学术的反叛,形成人文学科与社会科学的分野;那么,表2则意味着一个新的整体性学术时代的到来。表2还告诉我们,社会科学在第二个范式阶段形成的与人文学科的分野,借助于生成数据和一切皆可数据化,又重新回到了一致。素材的一致影响的自然是方法和理论的一致,尽管这些一致的出现还有一个过程,甚至是一个复杂的社会过程,机会却已经呈现,那就是人类又有了在更高维度对人文与社会重新进行整体性认识的机会,文科又有了重新整合的机会,这将是一个新的综合。

由数字化带来的文科范式革命,即智能革命。智能革命是分科化学科综合的开端,由于在数字化进程中大量使用机器,且机器承担的工作是人力难以承担的,具有不可替代性,因此,智能革命也是运用人机智能进行知识创造与积累的开始。这是人类历史上从未有过的格局,也是人类正在探索的知识创造与积累范式。

我们可以借用图1所示的范式迭代路径,对第三个文科范式的发展进行判断,这一进程大

<sup>①</sup> 当然,自然科学内部也有差异,物理学只是一个典型的例子。

致处在前范式阶段。在这个阶段,各种努力缤纷呈现,如阿尔法系列的努力、<sup>[11]</sup>计算社会科学的尝试、<sup>[12-14]</sup>科学研究人工智能化(AI for ...),<sup>[15]</sup>以及当下讨论热烈的 GPT<sup>[16]</sup>等。这些研究并非只是技术和方法的变革,而是由技术撬动的革命,实现了资源/素材、技术/方法与视角/理论之间的重新组合。这一轮范式革命在形式上是“一切皆可数据化”,本质上却是人机互生。人机互生意味着,人类的知识创造与积累不再是精英的特权,甚至也不再是学者的专属,而开始进入去个性化的、人机相互参与的技术与社会交互并进的过程,进而进入去个性化的世俗化过程。

在我们看来,这个过程至少需要经历四个阶段(见图 4)。

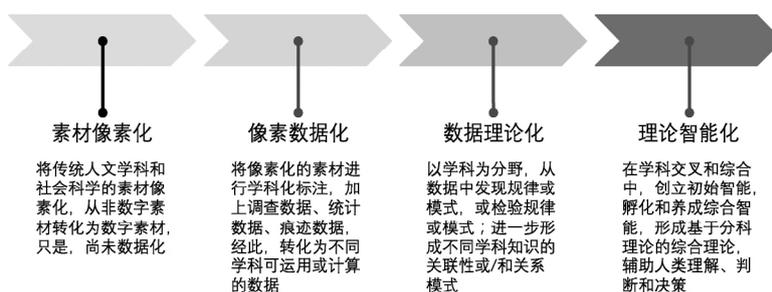


图 4 文科第三轮范式革命的四个阶段

### 三、第三轮文科范式革命的四个阶段

智能革命的前提是一切皆可数据化。数字技术进入大众化应用阶段以来,数据量的增长是一个显著现象。根据国际数据公司(IDC)的报告,基于连接泛在的数据量一直呈指数增长趋势,2016年约为20ZB,预估到2025年将达到175ZB。不过,这些数据主要是传感器和人类行为自动记录的数据,不包括非数字时代的知识积累数据。据估计,从直立行走到2013年,人类积累的可利用数据,大约只有5EB,只有2016年数据量的0.02%。尽管如此,它依然是我们认识人类过去唯一可用的数据,也是尚未完全数字化的数据。因此,将尚未数据化的素材数据化是智能革命优先需要完成的工作。

在实践中,这项工作又可以被划分为两个阶段,即素材像素化阶段和像素数据化阶段。第一阶段是将各种材质的素材转化为数字化版本,如图书、绘画、遗址、遗存、实物等的数字版本建设,如中国之外的古登堡计划、谷歌的数字化计划等,以及中国的古籍数字化、绘画数字化、遗址数字化、遗存数字化等。不过,在这个阶段,我们只是将各种材质的素材转化为可以用计算机存储的像素数据,将曾经局限于特定物理空间的、有各类限制的素材转变为可以通过计算机网络进行存储和传输的素材,却依然属于第一个范式和第二个范式的素材。

只有将像素化的素材数据化,才能用第三个范式的方法进行分析,这就是第二阶段,将数字化的像素素材数据化。纯粹的像素只是颜色的数值,也具有机器可识别和计算的意义。不过,机器只识别像素,却不识别像素承载的知识。如果希望机器也能识别像素承载的人类知识,如文本的社会含义、图画的人文精神等,就需要人将像素承载的人类知识告诉机器,让机器

有机会学习其中的含义、习得其中的含义、推演其中的含义。人指出像素承载的知识意涵的过程,被称为标注(labeling)。只有被标注了知识意涵的像素才是可以用于机器识别和计算进而进行知识创新与积累的数据,也只有数据化的像素素材才真正融入了数字化素材之中,为新的文科综合做准备。

标注是一个技术术语,指对像素化内容承载的人文与社会含义进行标记。例如,对一幅画承载的人物、事件、物品、山水、行动等进行标注,以至于仅仅透过标注就能反映画面试图表达的内容。一个简化的标注可以用人脸的特征点及其之间的关系来表示(见图5)。每个人的脸,其特征点位的参数、特征点位之间关系的参数都是不同的,对一张人脸的标注既有对特征点位的筛选,还有对特征点位之间关系刻画的建模。优化的标注应该满足如下标准:用尽可能少的特征点位和特征点位之间的关系最大限度地拟合出每一张脸之间的差异性。对包括文本在内的像素化素材进行标注,显然不是一个学科可以独立完成的,需要多个学科的协作,更需要机器尤其是人工智能的支撑,还需要社会化分工的共同努力。不仅各种材质的素材在像素化之后需要标注,生成的数据也需要标注。当前,数据标注已经成为一个庞大的产业,从知识创造与更新领域扩散到了社会生产领域。



图5 人脸标注

注:图片来自网络

赋予数据以意义,便让数据承载了信息,也让数据变成了可以从中发现或检验事物之间关系模式的信息化素材。文科第三轮范式革命的第三个阶段我们称为数据理论化阶段,即从数据中获得分科化的理论。图3的分科化,不仅有对象的细分,还有观察视角的细分。细分的优势在于可以从特定的视角观察和理解事物的属性以及事物之间的关系模式。一切属性都以数据形态呈现且携带标注,便为人类运用计算方法挖掘事物之间的关系模式提供了机会。以学科为分野的知识创造与积累便进入了一个新的阶段,既可以进行结构和格局的刻画,还可以进行动态与演化的刻画,因为数据不是静态的,而是流动的。在面对人类无力处理的数据量时,在面对动态与演化刻画的需求时,算法一跃成为支持人类知识创造与积累的第一技术与方法。因此,智能范式的基底是算法。

不过,一个学科的数据理论化还不是第三个范式的目标。第三个范式应该是以算法为底层,同时将第二个范式的人文学科与社会科学,甚至自然科学和工程,综合为基于人工智能支撑的、基于数据的理论检验和理论发现,输出的不再是人类试错式的知识,而是支持人类理解、判断、决策的,基于理论综合的科学依据,这就是我们认为是第四个阶段,即理论智能化。

鉴于第三轮范式革命的触发点是数字化,而数字化进程与人机互生密不可分,且人机互生的模式还在不断的、快速的迭代演进,因此,我们不能以为第三个范式的四个阶段是线性推进的。我们认为,一个概率更大的模式是四个阶段之间呈现为网络交互的演进,尤其是基于数据的标注、理论、智能之间的交互演进。一旦人工智能跨过奇点(singularity),接下来的格局会是什么,就很难预判了。自然语言大模型、生成式人工智能等向通用人工智能的快速发展,让我们很难人为预估第三个文科范式的具体演进路径。不过,目标应该是明确的,那就是基于人机互生的理论智能化。

因此,第三个文科范式的出现不意味着以分科范式为基准的库恩范式的失灵,而是需要我们以新的范式来理解文科的知识创造与积累。快速发展的语言学一反传统语言学对语意的阐释,从分析语料开始到对自然语言的拟合建模,形成千亿级参数的大模型,成为人工智能发展最前沿的领域之一,这便是新范式建设的一个案例。中国政府推动的新文科建设可以被理解为是新范式建设的组织化努力,各类数据库的建设可以被理解为第三个文科范式建设中第一阶段的工作。

#### 四、结 论

数字化的发展推动了人类对人文与社会认识的发展,触发了人类面向人文与社会的知识创造与积累的第三轮范式革命:智能革命。

第一轮范式革命发生在认知革命之后,伴随着书写语言和图画的出现,对理性的运用推动了人类思辨与逻辑的发展,以及对自然和社会的解释,让人类摆脱了对神谕的依赖、对超人类和超自然力量的信仰,促进人类对自然与社会的认识迈进了知识创造与积累的整体性学术阶段,形成了以知识精英为主体的学术范式。

第二轮范式革命发生在将数值型计量引入对人文和社会现象的测量和分析之后,即计量革命。统计思维和统计学的发展不仅为重新认识人文和社会素材提供了工具,还为分析数值型素材提供了方法和技术,构成对第一个文科范式的侵蚀,让对社会现象的刻画与分析形成了分离于整体性学术的分科知识,也推动了基于细分对象的分科性的知识创造与积累。第一个文科范式建立的整体性学术在分科学术的侵蚀下,被分解为人文学科和社会科学。第一个从整体性知识系统中分离出来的社会科学是经济学,随后是社会学、政治学等。计量革命还让从事知识创造与积累的人形成了一个共同体,且通过知识生产与知识运用的逻辑一致性体系,推动了个性化且世俗化的知识创造与积累的发展,形成了以知识应用为导向的学科范式。

第三轮范式革命发生在数字化进程中,即智能革命。触发第三轮范式革命的转折点在于数据的大量积累和一切皆可数据化,对数据内涵的人文与社会属性的挖掘以及基于数据的对事物之间关系模式的发现是智能范式的内核。从范式的成熟度判断,第三轮范式革命可能要经历四个阶段,当下正处在第二阶段早期。值得注意的是,与前两个范式不同,因为机器尤其是人工智能的介入,本轮范式革命的四个阶段呈现阶段之间交互迭代的格局,在奇点到来之

前,人类还在主导着面向人文与社会的知识创造与积累。智能革命引入了机器作为知识创造与积累的力量,形成了人机互生体,而机器的普遍使用既使得在第一、二个文科范式中一直存在的个性化知识创造与积累嬗变为去个性化的世俗化的过程,也使得之前的以分科为特点的知识创造与积累重新回到整体性的知识创造与积累状态。与第一个文科范式的整体性知识生产不同的是,这个阶段的整体性知识生产是以分科性知识创造与积累为前提的新综合。

给定人的生存具有了制度保障,智能范式再次让人类对人文与社会的认识回归于满足人类的精神诉求,在世俗化的知识的社会应用中,知识成为人类面向整体性生产和生活的真正的不可或缺的支撑。

### 参 考 文 献

- [1] HEY T, TANSLEY S, TOLLE K. The Fourth Paradigm; Data-intensive Scientific Discovery[R]. New York: Microsoft Research, 2009.
- [2] 托马斯·库恩. 科学革命的结构[M]. 金吾伦,胡新和,译. 北京:北京大学出版社,2012.
- [3] 保罗·费耶阿本德. 知识、科学与相对主义[M]. 陈健,等译. 南京:江苏人民出版社,2006.
- [4] 卡尔·波普尔. 猜想与反驳:科学知识的增长[M]. 傅季重,等译. 杭州:中国美术学院出版社,2003.
- [5] 伊姆雷·拉卡托斯. 科学研究纲领方法论[M]. 兰征,译. 上海:上海译文出版社,2005.
- [6] 尤瓦尔·赫拉利. 人类简史:从动物到上帝[M]. 林俊宏,译. 北京:中信出版集团,2017.
- [7] Miller George A. The Cognitive Revolution; A Historical Perspective[J]. Trends in Cognitive Sciences,2005,7(3): 141-144.
- [8] Michels Robert. Zur Soziologie des Parteiwesens in der modernen Demokratie; Untersuchungen über die oligarchischen Tendenzen des Gruppenlebens[R]. Leipzig, 1911.
- [9] BAYLY C A. The Birth of the Modern World, 1780-1914: Global Connections and Comparisons[M]. Medford, MA: John Wiley and Sons, 2018.
- [10] Richards Michael, Kate Britton. Archaeological Science: An Introduction[M]. New York: Cambridge University Press, 2020.
- [11] Callaway Ewen. What's Next for AlphaFold and the AI Protein-folding Revolution; DeepMind Software That Can Predict the 3D Shape of Proteins Is Already Changing Biology[J]. Nature, 2022, 604: 234-238.
- [12] Lazer David M J. Computational Social Science[J]. Science, 2009,323:721-723.
- [13] Lazer David M J. The Rise of the Social Algorithm[J]. Science,2015,348:1090-1099.
- [14] Lazer David M J, PENTLAND A, WATTS D J, et al. Computational Social Science: Obstacles and Opportunities[J]. Science, 2020,369(6507):1060-1062.
- [15] LEIST A K, KLEE M, KIM J H, et al. Mapping of Machine Learning Approaches for Description, Prediction, and Causal Inference in the Social and Health Sciences[J]. Science Advances, 2022,8(42): eabk1942.
- [16] Stokel Walker Chris, Van Noorden Richard. What ChatGPT and Generative AI Mean for Science[J]. Nature, 2023,614(7947): 214-216.

## Digitalization and Paradigm Revolution of Liberal Arts

QIU Zeqi

*(Department of Sociology, Peking University, Beijing 100871)*

**Abstract:** Digitalization, known as the third technological revolution experienced by humankind, is sweeping all corners of the world, and is also the driving factor of the ongoing third round of paradigm revolution in the liberal arts. The liberal arts paradigm refers to the logical consistency system from concept to practice, from teaching to scientific research, from academia to knowledge. It not only has the basis of Kuhn's paradigm, but also refers to the correlation and consistency system of resources/materials, methods/technologies, and theories/perspectives of the liberal arts. The first paradigm revolution in the liberal arts is an academic revolution, the second is a measurement revolution, and the third is an ongoing artificial intellectual revolution. The liberal arts empowered by digitalization and artificial intelligence that are being practiced in academia and businesses is an exploratory knowledge creation and accumulation from material pixelation to theoretical intelligence. The construction of the new liberal arts advocated and promoted by the state is an organizational practice of the liberal arts paradigm revolution in China. If the unfolding liberal arts paradigm revolution is divided into four stages (material pixelation/pixel data/data theory/theoretical intelligence), then the current efforts are actually still in the first and second stages, and the second, third, and fourth stages may be interactively progressing.

**Keywords:** digitalization; liberal arts paradigm; knowledge production

(责任编辑 张 丽)