

大数据时代的教育计算实验研究

余胜泉¹, 徐刘杰²

(1.北京师范大学 未来教育高精尖创新中心, 北京 100875;

2.北京师范大学 教育学部, 北京 100875)

[摘要] 大数据为教育研究带来新机遇,全样本的数据来源、因果分析向相关性研究的转变、计算机仿真方法的应用以及数据驱动实验的实现推动了教育研究范式转变。教育计算实验研究借鉴了计算社会学的思路,基于教育系统的客观事实,利用计算机仿真技术实现教育系统模拟,通过参数演化,自动化、快速生成大数据,基于大数据分析观察教育系统演化过程,总结教育演化规律,使教育研究可量化、可检验、可重复。基于大数据的计算实验为教育提供了新的研究方法与分析技术。基于大数据的计算实验研究在宏观层面可用于研究教育资源配置,在中观层面可以研究教育公共服务,在微观层面可以研究各要素及个体对教育系统的影响作用,帮助人们科学制定教育政策,改进教育系统。

[关键词] 大数据; 教育研究; 研究范式; 教育计算实验; 计算机仿真

[中图分类号] G434 **[文献标志码]** A

[作者简介] 余胜泉(1973—),男,江西鄱阳人。教授,主要从事移动教育与泛在学习、信息技术与课程整合、网络学习平台关键技术、区域性教育信息化等方面的研究。E-mail:yusq@bnu.edu.cn。

一、引言

教育研究范式分为思辨研究与实证研究,而实证研究又分为定量研究、定性研究和混合研究。思辨研究有着独特的本体论价值和突出的认识论价值,它关注解决“应然”问题,主要建构概念、理论和观点,并通过逻辑推理来解决概念的、规范的问题^[1]。但是,思辨研究以个体经验材料作为论据对论点进行论证,研究结论受个体理性认识能力和主观经验影响痕迹大,对个体的理性推演能力和理论功底要求也较高。而且,思辨研究过分注重对事物内部统一性的探求,从而容易忽视对事物内部的矛盾性或变动性的认识,容易出现用静止的观点看问题的倾向^[2]。

实证研究是基于事实和证据的研究,具有客观、量化、有定论、可检验等特征^[3]。教育实验研究是实证研究中最常用的一种方式,具有严格的实施规则,也是目前大多数研究者较易把握和使用的研究方法。但是,教育实验研究仍然存在一定的局限性。首先,教育

中与人的行为相关的情境比较复杂,许多干扰因素难以彻底排除,每个研究在时间、样本、情境方面都具有特殊性,一个独立的研究有时难以帮助人们在类似情境中基于该研究结果做出决策,由此导致研究结果在教育实践中的应用效果受到质疑^[4]。其次,人的生理特征、心理特征和社会特征的复杂性导致教育的数据化程度极其有限,无法保证教育实验研究在给定的完全相同的条件下进行,不可能真正控制各种内部与外部变量^[5],这就使得教育实验研究不同于自然科学实验,无法还原和重复。即便是相同的情境,相同的被试,由于人具有主观性,加上心理变化导致被试对干预手段的反应发生变化,致使研究结果发生改变,导致教育研究实验结果不可验证。因此,教育实验研究存在不可重复性、不可验证性、研究结果无法应用于现实等问题,导致教育研究的科学性无法确保。

随着数据密集型科学的快速发展,数据成为教育研究和利用的主要对象,大数据技术以“大数定律”为基础,通过相关性或关联性分析,寻找教育活动的内

部主导因素和高关联性外部因素,为教育决策、教育管理和教育研究提供技术支持,大数据技术引起了教育研究的范式转变。计算社会学使用计算实验方法,仿真模拟真实的社会系统,通过改变输入变量属性观察社会系统的演化过程和结果,总结社会系统的演化规律。大数据技术与计算社会学思路的结合,为教育研究提供了新的思路。

二、大数据教育研究新机遇

大数据具有数据量大、实时性强、涵盖面宽、结构形态多样、动态变化等特征⁶,为教育研究带来新的方法论和研究手段。

(一)教育研究的数据来源于全样本

虽然实证研究能够突破思辨性研究的局限,在一定程度上保证教育研究的科学性,但是在教育实证研究中,观察、实验和调查等方法获取的是小样本数据,是特定情境下的教育现象数据,是局部范围内的个体样本数据,是研究者主要关注的部分变量的数据,大部分属于结果性数据,而非过程性数据,由此得到的实证研究结果并不能科学地代表复杂的教育系统,得出的结论也难以适用于整个教育系统。而且,抽样的数量和代表性会影响研究结果的质量。

过去通过收集小样本数据研究教育问题和教育现象的方法已不能满足现代复杂的教育系统发展和教育研究的需求,当前亟须来源于整个教育系统的大数据支撑教育研究。教育大数据不受样本选择的影响,可以处理和某个教育现象相关的所有变量和所有数据,并由此发现数据背后的教育规律。教育大数据包括教育系统大数据和系统要素的个体大数据,是全样本、全过程的数据。系统大数据针对教育系统的所有要素以及所有要素的全部个体,涵盖了教育系统的全体样本和全体变量。对系统大数据进行数据挖掘和学习分析所得到的研究结论能够覆盖所有教育现象和教育问题,能够面向整个教育系统,能够准确地揭示和反映教育规律。个体大数据针对教育系统要素的每一个个体,数据来源于个体教育活动的全过程,涉及与个体相关的所有变量。个体大数据记录的是每一个个体的发展全过程,是连续性数据而不是局部数据,对个体的大数据进行挖掘和分析不仅能够揭示个体特征,而且能够对个体进行微观分析,了解个体的发展历程,发现个体成长规律,预测个体发展中可能遇到的问题,从而针对个体实施有效的干预,以促进个体成长和发展。

(二)教育研究的数据具有丰富的多态性

传统教育研究的数据来源具有明确的目的和对

象,且多是通过调查、访谈、观察等获得数据,数据维度比较单一。而且传统的数据收集方法无法排除人们主观性以及概念的误解等因素造成的数据不真实、不客观等问题,从而降低了教育研究的信度和效度。基于大数据的教育研究不再局限于单一维度的数据采集与统计分析,而是收集教育现象中所有形态的数据,包括结构化和非结构化数据,然后对多模态的数据从多种维度出发做统计性分析,从而精确地找出事物之间的相关关系,为研究者提供新的研究视角。例如:使用多态性数据分析不仅能够研究与学习有直接相关的认知因素,还能够探索影响学习的非认知因素。

大数据的多态性及大数据技术对结构化数据、半结构化数据和非结构化数据的全面采集,突破了传统教育研究中只能通过收集结构化数据实施“观察现象—发现问题—分析原因—提出对策”的因果研究的范式,使得教育研究者能够对体量庞大、类型多样的数据集进行对比分析、交叉检验和聚类统计,它更加强调发现而非证实,强调数据集之间的相互关联,强调数据正负误差的相互抵消,能够更多地从发现多个数据集之间的相关关系中找到问题、探索规律和预测趋势⁷。因此,大数据使得教育研究在不受样本影响的情况下处理与某个现象相关的数据,通过对多模态的数据进行相关分析,能够帮助人们发现数据背后的规律,帮助人们高效、准确地预测未来,从而建构教育研究的科学意义。

(三)教育研究的关注点转向相关性研究

教育系统的复杂性、人的行为的复杂性、教育结构的复杂性、教育系统各要素间相互作用的复杂性等均导致教育系统中各要素之间具有复杂的因果关系。而传统的教育研究通过对局部现象的研究,并将研究结论推广到整个教育系统中,希望解决整个教育系统存在的问题,这种以点带面的推广范式存在很大的局限性和不足。大数据时代,教育研究不再仅仅探索教育现象复杂的、难以定论的因果关系,主要通过统计分析海量数据中的相关性来探求教育规律。

大数据研究不同于传统的逻辑推理研究,而是对大数据做统计性的搜索、比较、聚类、分类等分析归纳,关注数据的相关性⁸。基于大数据的教育研究范式侧重于对客观数据的挖掘和分析,能够帮助人们研究与分析教育现象、教育活动、教育结果等,探究教育系统中各要素之间的相关性,确定教育与个人因素、群体因素之间的相关性,发现教育系统与其他社会系统之间的相关性。大数据教育研究可以对教育现象和教育问题的相关研究作因果关系解释,帮助人们寻找教

育系统的内部主导因素,寻找与教育系统具有高关联性的外部社会因素,从而为推动教育改革寻找突破口,为教育决策、教育管理和教育研究提供技术支持,为教育改革和实施教育实践活动提供理论依据。而且,教育大数据分析所得的相关关系不存在模糊性、不确定性和易变性,可以通过重复实践和研究加以检验、校正与论证。

(四)教育研究采用计算机仿真方法

教育实验研究方法通过严格控制一种或几种变量来验证、探讨教育现象的因果关系。但是在现实情境中,影响教育的因素众多且错综复杂,使得教育实验研究与现实的教育活动存在很多不一致,研究结论在教育应用中往往产生偏差。而且,人作为教育研究的对象,具有不确定性、不可逆性、不可控制性,这对教育实验提出了更高的要求,教育实验不允许出现差错或者失败,尤其是大规模实验更是不允许失败,失败不仅会造成资金和资源的浪费,还会对学生造成伤害。由此可见,通过单因素、严格的变量控制来寻求教育的局部改变,强调对教育实验因果关系的解释,追求实验结果的精确性、纯粹性的教育研究范式已经越来越难以适应现实情境中的教育实践活动的需要^[9],而计算机仿真为教育研究带来新的契机。

计算机仿真是依据系统原理,建立描述实际系统结构或行为过程的模仿系统,即利用具有一定逻辑关系和数量关系的仿真模型,在计算机平台上进行实验研究,以获得对实际系统的正确认识,确定决策方案^[10]。在教育研究中,通过界定现实的教育系统要素之间的关系,建立起教育系统要素关系的数学模型,然后利用计算机仿真技术进行实验,并对仿真实验结果进行评估与分析,帮助人们了解教育系统运行规律,从而支持人们做出准确的教育决策。利用计算机仿真技术开展计算实验研究,不仅可以降低成本,而且能够避免在真实情景中开展教育实验存在的风险。此外,计算机仿真实验可以重复实施,不断迭代,生成大数据,使实验结果真实地反映教育现象,揭示教育规律。

三、计算社会学的启示

传统教育研究范式虽然注重理论推演与实证研究相结合,但是由于教育实验的不可重复性和不可检验性、教育数据的小样本特征,使得教育研究结果的推广与实践检验受到限制。大数据为教育研究提供了新机遇,但也带来了新挑战。全样本的大数据采集与分析对数据采集、数据存储和可视化分析技术提出了

更高的要求,增加了教育研究的难度。这就需要新的研究范式、研究方法和技术来弥补传统教育研究范式的弊端,解决大数据教育研究的缺陷。计算社会学能够捕捉、获取和处理有关人类行为与社会运转的大规模数据^[11],通过研究人、社会行为、社会系统以及与计算技术互动中的相互作用^[12],能够获得对人类的社会、经济和政治等更深刻和更准确的理解,为教育研究带来了新的启示。

计算社会学立足于客观事实,利用计算机仿真实验对社会现象进行分析,能够得到确切的发现或结论。计算社会学将社会系统涉及的所有主体纳入研究对象中,并从整体论出发研究各个主体之间的相关关系,通过改变主体行为属性值来观察主体和客体在互动的关系中重新建构客体的演化过程,该演化过程强调研究过程的互动性,强调在自然情境下通过计算实验、仿真模拟等技术手段理解被研究者的行为和思想^[13]。计算社会学使用计算机仿真实验,可以自动化、快速、长时间运行并生成大数据,通过数据的演化,帮助人们洞察现实和预测未来。将计算社会学的方法应用到教育研究中,通过设计教育情境,理清教育系统所有主体、变量之间的相互作用关系,构建数学模型,同时,利用计算机仿真平台实施社会计算实验,能够全面模拟教育演化过程,探寻教育规律。此外,基于计算机的社会计算实验可以改变各种变量的属性值以反复实验,实现教育实验的可重复、可验证、可检验、可量化,从而推进教育研究范式发展。

计算实验通过计算机仿真模拟教育系统,处理的是连续动态的教育过程,生成的是连续型大数据,这既能保证数据的丰富和全面,又减少了由人的主观性判断带来的误差,帮助人们精准提炼教育规律。将计算实验应用到教育研究中能够穷尽所有变量的所有数据,演示教育系统演化的全部可能性,有助于确定教育系统在何种状态下、在哪些行为属性值下会出现瓶颈问题,并明晰在何种状态下可以突破瓶颈问题,改变哪些属性能够解决问题。计算实验让人们能够提前预知可能出现的演化路径,并通过改变决策来引导产生最优的演化路径。

计算社会学通过仿真模拟真实的教育系统,将教育系统中的所有变量纳入仿真系统中,模拟多元变量之间的相互作用,可视化分析教育现象的变化规律。计算实验能够将人的认知、心理、态度、行为等主观和客观因素作为变量,将自然环境、社会环境作为变量,通过建立人、自然环境、社会环境等因素之间的数学模型,在计算仿真实验中同时考虑多元变量及其相互

关系。通过对教育系统内部各因素之间的关联性、教育与社会系统外部因素之间的关联性进行模拟,演示教育系统中各要素之间的作用关系,多元变量作用下的教育演化规律,如学生的成长规律、人类的发展规律、教学规律、教育决策执行规律等,可以验证教育系统各要素行为的有效性,验证教育决策的有效性,评估教育决策的效益,为预测教育决策带来的教育结果提供数据支持。

四、基于大数据的计算实验研究

大数据为人们研究社会科学现象的变化以及研究社会科学现象中的高度关联和影响作用创造了新的机会,能够交叉融合社会行为和数据计算处理系统,用以研究复杂社会系统运行的规律与发展趋势,带来了社会科学研究范式的转变^[14]。现代社会条件下的教育系统涉及自然、社会、经济、技术等各个领域,具有系统性、复杂性特点,因此,教育系统被认为是非线性的、复杂的大系统。根据已有数据建立计算仿真模型,进行可重复、可复原、可验证的“仿真计算实验”,通过计算实验生成教育大数据,从整体论、系统论出发进行情景分析与政策模拟,分析教育现象和教育问题,是解决教育系统治理的一个关键技术和发展趋势。计算社会学与大数据技术的结合,为教育研究提供了方法论基础、技术基础和实验研究基础,基于计算社会学思想,运用大数据技术的计算实验研究可以成为教育研究的新范式。

(一) 基于大数据的教育计算实验研究方法

在教育研究中,数学模型的介入使教育系统中各要素的关系能够量化处理,这为应用计算机技术实现仿真模拟研究提供了技术和方法支持。计算实验研究利用计算机仿真模拟现实教育系统的演化规律和教育结果,在建立教育系统各要素关系的数学模型的基础上,通过计算实验仿真平台实施实验,通过设定不同的输入来产生不同场景的数据,通过数据分析来观察输出,为教育决策提供支持。

基于大数据的教育计算实验研究是采用定性与定量相结合的方法进行问题的研究。用系统动力学理论和方法分析真实的教育系统,利用教育计算方法,建立概念与定量一体化的模型,并借助计算机模拟技术对教育问题进行各种条件下的演化,模拟教育系统的演变过程和行为,模拟决策对象或决策过程发展变化的规律,并根据模拟出的定性与定量结果调整政策以获得最佳决策效果。计算实验研究在非线性的复杂教育系统的仿真研究中体现出了优越性。

基于大数据的教育计算实验研究是数据驱动的实验研究,将现实中教育系统的各要素及其关系映射到仿真模型,将研究问题所涉及的显变量、潜变量及其相互作用关系一一表达出来,建立基于计算机的人工仿真教育系统,这个系统是真实教育系统的镜像。因此,可以将数据驱动的教育实验表示为:数据驱动的教育实验=真实教育系统+仿真教育系统+数据分析引擎,它们之间的关系如图1所示。



图1 数据驱动的教育计算实验研究各要素间的关系

基于大数据的教育计算实验研究用虚拟现实、决策仿真对教育系统进行建模,抽取真实教育系统中的运行要素,梳理要素之间的关系与规则,建立虚拟的仿真教育系统。同时,在进化的思想指导下,对仿真教育系统进行演化,并根据真实的教育情境设定所有自变量参数,使自变量参数从极小值向极大值演化,实现全样本数据输入,输出的是教育系统在每一个样本变量值下的演化过程和结果,据此能够穷尽所有教育现象。在计算机仿真模拟决策支持的整个过程中,人们可以通过对仿真模拟系统运行状态的跟踪和分析来研究真实教育系统及其运行过程,通过模拟教育系统运行状态的各项参数及其变化趋势,以此来预测和推断实际教育系统对象或教育系统的真实演化规律,从而为人们的科学决策提供真实有效的数据和依据。此外,针对教育系统演化过程和结果数据,对其进行深度挖掘与解析,可视化展示数据分析结果,能够有效帮助人们直观理解教育现象,抓住教育问题的本质,可以发现如何改进教育系统的具体要素和变量以解决教育问题,从而实现教育系统的最大价值。参照仿真系统的逻辑调节教育系统要素的关系,改变教育系统要素的属性,从而指导真实的教育实践活动,优化教育系统。

基于大数据的教育计算实验通过控制教育系统的变量,促使大数据的产生,探索教育现象的演化规律,属于数据驱动实验。数据驱动实验的特征表现在三个方面:一是控制变量不再是传统实验中的一个或若干个变量,而是将所有可能的变量都考虑进去,实现控制变量的全样本考虑,并根据教育系统各要素之间的相互作用关系设计数学模型,从而能够真实反映现实教育系统的复杂性。二是控制变量的取值不再是传统实验中的一个值或者若干个值,也不是若干个以

等级表示的模糊值,而是变量所有可能的取值,通过计算机自动产生大数据。控制变量的大数据输入为教育实验提供了所有可能性的条件,改变了传统教育实验中只输入能够获得显著效果的变量值,实现能够观察在正常状态下和极端条件下的教育现象变化,这种变化不仅涵盖教育实验的显著性效果,还包括所产生的不良后果,从而使得教育实验能够真实反映教育规律。三是数据驱动实验不仅关注输入和输出,还强调实验过程中各种信息的数字化、数据化,记录实验过程中生成的大数据,以此分析变量变化过程中相互之间的作用机制,打开教育现象变化的“黑箱”,使得教育演化过程透明化。

(二) 基于大数据的教育计算实验研究框架

基于大数据的教育计算实验研究利用仿真系统和演化引擎对实验数据进行挖掘和分析,以研究计算实验过程中教育系统的演化规律,提出改进真实教育系统的干预政策和方法,其框架如图2所示,核心要素包括教育系统的模拟仿真、影响教育系统关键要素的参数化及其演化、度量效果的教育价值计算与趋势识别。

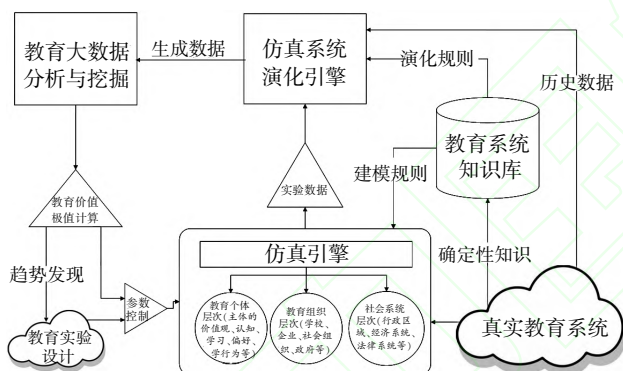


图2 基于大数据的教育计算实验研究框架

1. 教育复杂系统的计算仿真

教育计算实验研究首先要针对复杂的教育系统进行系统分析,在此基础上建立仿真系统。教育系统分析的流程主要包括:(1)社会、经济、教育、技术系统与环境的划分;(2)系统内各子系统及其层次的划分,基于真实的教育系统进行定性模型框架的建立;(3)关键利益相关者及其行为的定义,关键参与对象之间的互动模式及其作用关系;(4)宏观、微观要素及指标体系的建立、系统间相互作用的研究;(5)基于已有学校的数据进行各种定量模型的建立及系统参数的拟合;(6)定量模型结果的解释与预测以及系统政策参数对结果的影响研究等。其次是开发教育仿真系统。教育仿真系统的各要素及其关系、规则来源于真实教育系统和教育系统知识库。从真实教育系统中提取教育要素及其行为规则;从教育系统知识库中提取各要素及其作业规

则,从而建立教育个体层次、教育组织层次和教育系统层次等不同层面融合的仿真引擎。最后,基于仿真引擎,利用真实教育系统中的历史数据,提取教育系统知识库中的演化规则,通过输入实验数据(数据学习),利用仿真系统演化引擎实施教育系统的仿真演化。

2. 系统演化、大数据的生成与深度挖掘

教育仿真系统演化的实施依赖于仿真引擎对历史数据的学习,对教育要素及其规则的提取与设计。根据教育系统内的主体要素,仿真系统演化引擎从历史数据中提取相关主体要素的行为模式及其行为规则,构建教育系统的仿真演化规则,在仿真系统演化规则控制下,任何变量的属性变化所引起的教育系统变化都会被记录并生成实验数据。而且在教育系统演化预测中引进智能演化算法(如遗传算法),可以使教育系统演化计算中教育主体的行为不再是事先给定的固定数据,而是在自组织和广义进化机制作用下,主体根据外部环境和自身属性适应性地调整行为规则,并在对系统状态判断的基础上生成的动态数据,以实现计算机自动生成输入数据,从而实现全样本数据的收集。演化只生成原始数据,数据的规律、意义及其解释需要进一步通过深度挖掘来处理,并通过可视化的方式展示出来,让决策者以直观形象的方式理解复杂变量影响下的教育系统运行规律。利用仿真系统实施动态演化,生成教育大数据,从而对数据进行可视化分析和深度挖掘,一方面把教育系统的宏观状态和每个个体在教育行为层面的微观决策联系起来,另一方面在演化计算中,从微观层面上模拟了教育系统的动力机制。

3. 演化趋势及教育价值的极值计算

在教育计算实验研究中,教育价值是实验设计中期望效果的量化表示。仿真模拟教育系统的演化趋势和计算教育价值的变化趋势可用于指导并预测真实教育系统的发展趋势。通过大数据分析 with 挖掘以探索教育各要素之间的作用关系,通过改变教育要素的属性值和利用仿真系统观察教育价值的变化趋势,从而识别出教育价值的极值及引起极值变化的教育要素属性值。此外,根据教育仿真系统中教育价值的极值计算结果可以指导真实教育系统的发展,并有效引导良性教育价值的极值发展,同时,规避恶性极值。另外,根据教育系统的计算实验结果做聚类、相关、关联等数据挖掘,可以提炼出影响教育系统发展的关键因素并对其进行解释,还可以以可视化的方式理解关键要素相互作用的传导链条。最后,对计算实验结果的讨论解释,有利于产生应对真实教育系统发展变化的干预措施,有助于对真实教育系统进行预警和干预,为教育决策

提供依据。同时,对教育系统演化中出现的特殊现象进行解释和讨论,有助于发展和形成新的教育科学理论。

4. 真实教育系统的改进实践与检验

通过教育仿真系统的计算实验与演化能够发现影响教育系统发展的关键因素及其影响程度。将教育仿真系统产生的最优解、教育价值最大化的极值、教育仿真系统演化过程中发现的规律等用于设计教育实验,应用到真实的教育场景中,改进真实的教育系统,同时,实践还可以检验教育仿真系统的信效度和科学性。教育计算实验发现的规律,再结合人类的主观经验和洞察力,可以实现人机结合的群体决策,在仿真演化的基础上,综合集成专家的专业知识、经验以及各种模型运行结果,引入人机智能对话系统,基于专家的经验性知识和模型的模拟力量,可以充分展现假想情境下的系统演化,帮助人们进行情境分析与策略选择,并为制定相应的、主动的教育政策或被动的教育事务应急方案提供科学的决策依据。同时,在教育实践中产生教育大数据,发现新的变量与教育系统演化规则,并反过来用以改善教育仿真系统,实现教育仿真系统的迭代发展。通过对改进的教育仿真系统进行新一轮的演化,生成新的教育系统演化状态最优时各影响变量的值,以及当教育价值达到最优时各影响变量的极值,从而生成新的教育决策,并将其应用到真实的教育系统中,帮助实施新一轮的教育系统改进,不断迭代完善真实教育系统。

五、基于大数据的计算实验研究层次

计算实验应用于教育研究,使得教育实验研究可以综合考虑多种因素,使得教育实验更加接近现实情境,使得教育实验可还原、可重复、可检验、可验证。根据教育系统涉及的范围、范畴,可对计算实验的研究层次进行划分,宏观上可研究教育资源配置,中观上可研究教育公共服务,微观上可研究教育政策和策略的实施过程及各教育要素的演化过程,从而分析教育政策和策略的实施效果,为改进教育政策和策略奠定基础。

(一)宏观:教育资源配置

教育资源不仅包括教育信息资源和人力资源,还包括教育经济资源和教育政策资源。教育资源配置就是对教育信息资源、人力资源、教育经济资源和教育政策资源等进行合理的分配。在教育资源配置研究方面,更多的研究是从理论演绎和思辨出发分析和构建教育资源配置方法、机制;还有一部分研究是通过局部调查,从实证角度分析教育资源配置现状,从而发

现问题并提出解决方案;也有研究通过对现实的教育系统进行关系分析,利用系统动力学方法、数学建模方法等分析教育资源配置的效益。但是,思辨性研究并不能真实反映现实教育资源配置现状;基于小样本的实证研究不能发现教育资源配置的全局特征,静态的截面数据不能发现教育资源配置的动态特征,数据演化不能预测教育资源配置相关要素的变化;系统动力学和数学建模方法是在获取现实数据基础之上进行的仿真模拟,其只能分析主要的教育系统要素,难以涵盖所有教育要素,尤其是对与人有关的社会属性(认知水平、心理状态等)方面难以量化处理。

教育资源配置涉及教育系统中政府、学校、社会、学生、教师、家长等多个主体,受到社会经济发展、地理环境、社会文化等因素影响。基于大数据的计算实验能够从整体上分析教育资源配置涉及的所有要素,全面把握教育要素之间的相互关系,分析教育资源配置的实施对教育系统、教育公平的作用,分析不同资源配置方案对提升教育生产力的关系。通过参数设置、系统演化可以观察不同资源配置下的教育系统的演化规律。计算实验研究能够分析多主体之间的博弈关系,能够研究实现教育资源配置效益提升的主体属性,能够在满足各个主体均衡的、最优的收益基础上实现教育资源配置的效益最大化。计算实验利用计算机仿真平台实施可重复性的实验研究,将过去排除在教育决策之外的海量的、异构的数据纳入教育问题的决策过程中,形成教育大数据,找到教育资源配置的关键点与盲点。通过相关分析、社会网络分析、高维度数据分析、聚类分析、关系挖掘等关联性和深度化的分析与挖掘,可以洞悉教育行为和现象背后所蕴含的深层逻辑,并对特定资源配置条件下未来的发展进行有效预测,进而为解决现实的教育治理难题、促进教育资源配置的科学化和治理过程的精细化提供全新的技术支撑。

(二)中观:教育公共服务

新时代的教育矛盾是人们日益增长的高质量教育需求与教育质量不平衡、不充分发展之间的矛盾,表现在教育公共服务需求上就是人们对教育公共服务的质量公平需求逐渐增强。如何实现教育公共服务质量公平供给是需要解决的难题,在研究教育公共服务理论上,加强对服务供给主体、供给方式、供给效益的实证研究是非常重要的,它将直接影响教育公共服务供给实践,影响通过分析供给效果、效率和效益得到的政策的有效性和评价的科学性。由于研究方法、研究手段和技术工具的限制,传统方式难以做到大规模的实验研究,教育公共服务供给的政策实施也大多是在

理论研究基础上实施的教育公共服务试验,或者是在小范围内试验成功后的大面积推广,这显然没有考虑到地区差异、文化差异和人的差异。一旦大规模实施出现问题,造成的损失就比较大且难以弥补。

教育作为公共服务最具影响力的领域之一,势必要突破传统的仅提供有限的管理和服务范畴,教育管理者的决策也不能仅仅依靠直观经验和简单的统计数据。将基于大数据的计算实验应用于教育公共服务供给实验研究中,通过模拟教育公共服务供给系统,改变各种变量和属性来观察教育公共服务供给的效果、效率和效益,同时,洞察有效的教育公共服务的结构,从而穷尽公共服务供给过程中存在的所有可能性,为人们应对和解决教育公共服务供给可能遇到的问题提前做好准备和支持。在区域教育层面,基于多元数据融合,利用数据挖掘技术和空间分析方法,为教育管理者提供全面客观的教育问题分析,并通过智能算法为区域教育供给侧改革提供决策支撑模型,促进“基于数据说话”教育公共服务方式和“动态实时”教育公共服务模式的实现^[15],辅助提高管理与服务效能,提升服务质量与获得感。

(三)微观:教育改进策略

要设计微观上解决各类教育实践问题的解决方案,就需要进行教育实验研究,要在教育实施场景中研究教育的运行过程,分析教育系统的运行规律,从实证角度分析教育政策的价值和结果,获得对教育政策的实然性认识,为决策者改进教育过程提供依据。由于传统教育实验研究对象主要是学生,不能做长时间的实验对象,以免实验对学生产生不良后果,影响学生身心发展,因此,部分对人的发展有高利害的实验应该结合计算仿真来实施,以预测可能产生的不良效果,避免教育实验对学生造成持久的负面影响。基于大数据的计算实验能够仿真模拟教育实践活动,模拟干预措施的实施过程,使得研究者能够全面观察教育干预措施的实施,并且能够避免现实中产生的不良后果。基于大数据的计算实验能够穷尽教育干预措施实施中出现的所有现象、问题和困难,观察它们产生的因素及其属性,借此帮助人们预测真实环境下教育干预措施实施中可能会出现的问题和困难,为决策者改进教育策略,完善教育干预措施提供数据支撑。

教育计算实验研究通过利用计算机进行仿真模拟真实的教育系统,以观察教育系统演化状态和发展趋势,探究影响教育系统演化的关键因素,发现教育系统演化规律、推演控制和影响教育系统发展的干预措施,而将这些影响因素、教育系统发展规律、教育干

预措施应用于真实的教育系统中,可以推进教育系统良性发展。根据教育系统演化中产生最优教育价值和最佳教育演化状态时各个教育系统影响因素的取值,设计真实教育系统中相应影响因素的取值,制定教育干预措施,以指导和推进真实教育系统的发展。

六、大数据时代教育研究的第四范式

从研究方法角度看,科学研究范式主要包括四类:一是经验科学研究范式,偏重于经验事实的描述,主要的研究方法是对自然现象的描述论证,对自然现象的系统归类;二是理论研究范式,偏重于理论总结和理性概括,强调较高普遍的理论认识,主要采用建模方式,由特殊到一般进行推演;三是计算仿真研究范式(第三范式),偏重于数据模型构建,利用定量分析及计算机来分析和解决科学问题,主要方法为利用计算方式模拟复杂现象,其中科学数据通过模拟的方法获得,而不再依赖于单一的实验;四是数据密集型科学研究范式,又称为第四范式^[16],主要利用计算机、数据管理和统计分析工具对大数据进行分析,发现复杂现象背后隐含的模式,并从中获取知识。

由于教育研究的对象主要是人,教育系统具有复杂性,研究活动与研究对象(被试)具有高利害关系等特征,且教育实验难以重复和检验,这就使得单纯依靠真实教育场景中的实验研究难以真正揭示教育系统的本质。此外,通过教育实验研究获取数据的操作非常复杂,难度大,采集的数据量小,一般是小规模的教育实验研究,很难做到全样本的数据采集,由此很难探索教育发展规律。这种情况下,数据密集型研究范式难以实施。计算仿真研究能够解决数据量小和数据覆盖面有限的问题,但是由于教育计算仿真研究范式无法真实还原教育系统,需要对真实教育系统做大量调研,模型拟合,检验和验证,迭代改进系统,即计算仿真研究必须在大数据结合下才具有存在的意义与价值,对教育研究才能够发挥其应有的作用。

因此,需要利用计算实验仿真研究,通过数据学习拟合教育系统模型,通过计算仿真演化生成教育大数据,然后进行数据挖掘和分析,发现影响教育系统演化的关键因素,提炼教育系统演化规律。同时,将计算实验仿真结果应用到真实教育系统中进行检验和验证,改进教育系统,并生成新的教育数据;再次拟合教育系统模型,可以进行新一轮的教育仿真实验,探究新的影响因素,改善或发现新的教育规律。这种“教育实践—数据生成—数据学习—模型拟合—仿真演化—教育实践”的迭代循环以及教育仿真系统的迭代

改进,就是将第三范式与第四范式相结合,从计算实验角度分析教育问题,挖掘教育数据。因此,教育科学研究应将第三范式与第四范式加以融合,形成具有教育计算实验形态的第四研究范式。

[参考文献]

- [1] 姚计海. 教育实证研究方法的范式问题与反思[J]. 华东师范大学学报(教育科学版), 2017, 35(3): 64-71, 169-170.
- [2] 彭荣础. 思辨研究方法: 历史、困境与前景[J]. 大学教育科学, 2011(5): 86-88.
- [3] 袁振国. 实证研究是教育学走向科学的必要途径[J]. 华东师范大学学报(教育科学版), 2017(3): 4-17, 168.
- [4] ÜSTÜN U, ERYILMAZ A. A research methodology to conduct effective research syntheses; meta-analysis [J]. Education and science, 2014, 39(174): 1-32.
- [5] 毋小勇, 张卫民. 理性面对教育研究与教育决策的“实证科学”化[J]. 教育发展研究, 2017(17): 31-37.
- [6] 余胜泉, 李晓庆. 基于大数据的区域教育质量分析与改进研究[J]. 电化教育研究, 2017, 38(7): 5-12.
- [7] 邹太龙, 易连云. 从“始于假设”到“基于数据”——大数据时代教育研究范式的转型[J]. 教育研究与实验, 2017(4): 74-79.
- [8] 李国杰, 程学旗. 大数据研究: 未来科技及经济社会发展的重大战略领域——大数据的研究现状与科学思考[J]. 中国科学院院刊, 2012(6): 647-657.
- [9] 杨莉娟, 郝志军. 教育实验研究的新思路·新探索·新成果——中国教育学会教育实验研究分会 2015 年学术年会综述[J]. 教育研究与实验, 2015(6): 95-96.
- [10] 盛昭瀚, 张军, 杜建国, 等. 社会科学计算实验理论与应用[M]. 上海: 上海三联书店, 2009.
- [11] 王国成. 行为大数据, 通宏洞微与人类决策——计算社会科学的兴起与发展[J]. 科研信息化技术与应用, 2015(6): 3-11.
- [12] DRYER D C, EISBACH C, ARK W S. At what cost pervasive? A social computing view of mobile computing systems[J]. IBM systems journal, 1999, 38(4): 652-676.
- [13] 王飞跃. 社会计算——科学、技术与人文的数字化动态交融[J]. 中国基础科学, 2005(5): 5-12.
- [14] LAZER D, PENTLAND A, ADAMIC I, et al. Computational social science [J]. Science, 2009, 323(1): 721-723.
- [15] 余胜泉, 汪晓凤. “互联网+”时代的教育供给转型与变革[J]. 开放教育研究, 2017, 23(1): 29-36.
- [16] HEY T, TANSLEY S, TOLLE K. The fourth paradigm: data-intensive scientific discovery [M]. Redmond, Washington: Microsoft Research, 2009.

Education Computational Experiment Research in the Era of Big Data

YU Shengquan¹, XU Liuji²

(1. Advanced Innovation Center for Future Education, Beijing Normal University, Beijing 100875;

2. Faculty of Education, Beijing Normal University, Beijing 100875)

[Abstract] Big data brings new opportunities to education research. Data sources of all samples, the transition from causal analysis to correlation research, the application of computer simulation methods, and realization of data-driven experiment promote the shift in educational research paradigm. Based on computational sociology and the objective facts of education system, education computational experimental research could achieve education system simulation by using computer simulation technology. Through parameter evolution, big data could be generated automatically and quickly. Based on big data analysis, the evolution process of education system is observed, and the laws of evolution education are summarized to make education research quantifiable, verifiable and repeatable. The computational experiment based on big data provides a new research method and analysis technology for education. Computational experimental research based on big data can be used to study the allocation of educational resources at the macro level, to study public service of education at the medium level, to study the impact of various factors and individuals on education system at the micro level, so as to help people formulate education policies scientifically and improve education system.

[Keywords] Big Data; Education Research; Research Paradigm; Education Computational Experiment; Computer Simulation