

面向智慧教育的大数据研究与实践： 价值发现与路径探索

肖玉敏，孟冰纹，唐婷婷，吴永和，谢 蓉

(华东师范大学 教育学部，上海 200062)

[摘 要] 大数据是越来越多的组织和个人努力去理解并利用其优势的一种存在。和其他领域一样，教育大数据能够帮助我们加深对学习者成长的理解，基于数据做出对个人和群体发展的预测，最终实现对人性的洞察，提供更符合人的需求的教育。文章探讨了智慧教育与教育大数据研究的关系，阐明教育大数据研究的价值体现在拓宽对教育的认知、实现基于数据的教育决策、促进教育的进化和培育大数据文化等四个方面，绘制了教育大数据应用研究的路线图。

[关键词] 教育大数据；智慧教育；基于数据的决策；路线图

[中图分类号] G434

[文献标志码] A

[作者简介] 肖玉敏(1968—)，女，四川成都人。副研究馆员，博士，主要从事数字化环境中的学习研究。E-mail: ymxiao@ses.ecnu.edu.cn。

一、引 言

随着大数据时代，特别是“互联网+”的到来，“教育领域正在开展的一些教育实验，如微课程、翻转课堂、慕课、电子书包等，都是寻求教育变革的一些尝试。如果仔细分析这些实验，不难发现它们都离不开数据。类似的教学改革都围绕为学生提供学习资源，并在学生学习过程中获取学习数据，随后积累、分析和应用这些数据，然后为他们提供评价、反馈、学习内容和交流互动的机会，提升学习质量。也就是说，基于数据的教与学实际上是当前众多教育变革的方向”^[1]。

然而从目前的实验我们也能看到，已有的数据之间的联系并没有建立起来，学校、教师、教育企业所拥有的学生学习直接相关的数据（一些个人基本信息数据在区域层面应该已经能够共享）基本处于各自利用

或者仅仅是存储状态^①，数据之间缺乏共享、交流和相互印证，因此从大数据的特征来看，其价值性（Value）恐怕会打折扣。

从教学的视角来看，虽然我们的学生已经充分利用数字化环境开展多种形式的学习，但是现有的学习分析系统基本上采集的还是他们在完成各种练习时的数据，并不能全面反映学生的真实学习状态。教师一般看到的和关心的数据主要也是学业相关的数据。从目前采集的情况看，数据的类型还相对集中在学生学习过程中点击、浏览和上传作业的数据，得出的结论基本上是为知识点查漏补缺服务，因此在数据的多样性（Variety）方面还相对不足。

此外，虽然在线教育和一些学校创新的教育实验帮助我们获得了大量的教育数据，但是数据量的增大并不一定保证能够给我们带来对教育洞察的结果，一

基金项目：全国教育科学“十二五”规划2014年度国家一般课题“智慧教育环境的构建与应用研究”（课题编号：BCA140051）；上海市浦江人才计划项目“基于教育大数据的学习分析教育应用创新研究”（项目编号：14PJC034）

①笔者查阅了大量文献，如《中国基础教育大数据》《大数据时代的教育》《教育大数据》以及相关的一些论文，发现大部分国内的案例主要体现的是区域或者学校的某个系统数据收集、分析的案例，除国家政策相关的数据系统，大部分基础教育的数据分析、应用都局限于小的范围。这和国外的illuminate系统、Knewton系统以及可汗学院等大不相同。

些非常具体的教育问题,如择校问题、学生的差异性发展问题、成长环境对学生的影响问题等,影响它们的因素千千万万,它们的分析、整理、相关性求证、得出可靠的结论等除了需要大数据分析工具的参与,更需要人的智慧。对教育大数据而言,其复杂性不容小觑。

可见,教育大数据研究将是一项需要更多数据积累与分享、更多类型数据保证其多样性和全面性、严峻考验人类智慧的研究,它是帮助我们理解教育相关的各个要素的重要途径。因此,我们需要深入地考察其对教育的价值,思考和明确其研究路径。

二、各行业大数据已有研究及启示

教育领域大数据的研究与实践离不开社会大环境的影响和支持,因此我们在聚焦教育领域大数据之前,应先了解大数据是怎样影响其他领域的。

(一)各行业大数据已有研究与实践

大数据是越来越多的组织和个人努力去理解并利用其优势的一种存在。以众多企业为例,它们都在不断地学习利用大数据带来的优势,通过传感器、射频识别仪器等理解商业环境,设计新的产品和服务,带着高期望,花费巨额资金,相信大数据是在信息时代保持高速增长的新的发动机,认识到企业的竞争力如何,也将越来越取决于是否有能力充分利用新技术的优势,而这些技术与大数据密切相关^[2],利用大数据进行产品或服务创新、获得全方位收益的企业成功案例不断见诸于各种报道和报告,如亚马逊公司利用用户浏览与购买的大数据进行精准的购物推荐,计算机科学公司(CSC, Computer Sciences Corp)利用大数据搜索为癌症患者提供治疗需要的精准干细胞配型,商业或投资银行基于大数据分析实现精准客户服务、提升用户体验、提高运营效率、减少投资风险、加强员工合作等。从这些案例和关于大数据的研究中,我们看到的不仅仅是数据、数据规模和数据分析与测量工具,而是一个更复杂的系统,在这个系统里,技术、数据处理和人都是重要的组成部分。在数据搜集、管理和利用的所有环节,技术发挥着重要的作用,服务于数据的整合、聚合和校准(Alignment),帮助人们进行数据分析,实现数据可视化,达到充分利用数据优势进行有效决策的目的;数据处理是大数据能够被有效利用的必备环节,在大数据领域通常被称之为数据治理。数据治理意味着规则、决策权和问责。对任何组织而言,通过数据治理可以了解造成某些数据质量较差的原因以及它们会带来的后果,这样就可以提前预测和纠正可能发生的问题,规划可行问题解决方案;在

大数据时代,数据显然是一个组织的财富,而数据的搜集、分析和使用都离不开人的参与,基于数据的决策也是由人来做出的,所以数据科学家的作用就尤为重要^[3]。

如果超越企业对大数据的解读与应用,从社会文化的视角来看大数据,大数据本身带来了一种新型的数据文化,这种文化提醒我们,随着互联、智能化技术的广泛使用,人们自觉自愿地贡献着各种各样巨量的数据,在 Hadoop 等分布式系统基础架构的帮助下,这些数据帮助人们实现对社会纷繁复杂事务的关注、分享、交流,信息传播速度快、范围广且不可逆转,因此,整个人类社会生活既体现出欣欣向荣,又表现出危机四伏,如患病既有更多的机会得到救助、捐助,也更有可能在生活中被嫌弃与排斥;大数据既营造了开放、共享的数字化文化氛围,为人类智慧的传播、文明的进化创造了过去任何时代都难以企及的机会,但是部分公司对数据的垄断又意味着人们的隐私无处可逃,其后果是如果信息泄露,将带来很多麻烦,每个人都不可能独善其身。当人工智能、虚拟现实技术更为普及,数据存储更加方便,技术所能处理的数据和数据形态更为广阔与复杂的情况下,大数据到底为人类文明带来怎样的未来,现在还难以预测。因此,如何对大数据带来的社会文化现象做合情合理的分析、诠释或者引导,需十分重视。

总之,各行业的大数据研究寻求大数据所带来的洞察力、价值和未来前景,普遍追求通过大数据获得商业智能(Business Intelligence),这种智能的过程是“将数据转化为信息,然后以此为基础制定决策,最终开始行动”^[4]。从企业对大数据的实际应用来看,各个行业领域对大数据虽然有不同的应用方式,但是基于行业特点有一些基础性的共识,大数据研究的主要目标在于:(1)洞察以获得远见卓识,挖掘商机;(2)提供精准服务改善用户体验,以便在竞争中处于优势地位^[2];(3)创造新的基于数据决策的企业文化。而从社会文化发展来说,大数据对于今天的我们是机遇与挑战并存的。如何充分利用其优势,及时洞察社会问题,厘清隐私边界,促进文化交流与社会发展,将是大数据研究与实践的重要主题。

(二)对教育大数据的启示

教育行业虽然与其他行业不完全一样,但是其他行业的大数据应用可为教育行业提供一些线索和值得借鉴的经验。具体表现在以下几个方面:

首先,在正确的支持理念下应用大数据。无论哪个领域,大数据的应用都是在一定的理念支持下开展

的。如凯撒医疗集团提出的医疗大数据的价值观框架是正确的生活、正确的关爱、正确的供给、正确的价值观和正确的创新。^[5]对教育领域而言,我们要帮助学生正确地成长,也需要基本的教育理念支撑。在智能时代,学生需要为未来不存在的工作和现在还难以想象的环境做好准备,因此,培养他们的自主学习能力、协作能力等终身学习能力就变得非常重要,所以,以学生为中心的教与学应该成为我们助力学生成长的基本理念。教育大数据的使命也将是支持以学生为中心的教与学。

其次,数据的获取渠道应该更多元化和更畅通。数据的采集与获取,教育领域目前基本上还是以在线课程的数据为主,在一定程度上还是相当有局限的。各行业的大数据应用环境,对我们也有启发。如何通过更多的利用技术,一方面提升学生的学习动机和学习效果,另一方面支持学生在学习过程中不断自然地上传数据,也是教育信息化努力的方向。

第三,大数据分析技术借鉴其他行业领域的基本框架。本文不探讨技术问题,因此没有阐述其他行业大数据分析的方法和过程。但是其他行业在技术上所实施的的分析模型对教育领域也有启发。英国学者Bart Baesens在《大数据分析:数据科学应用场景与实践精髓》一书中,对大数据分析过程模型做了比较清晰的解释:第一步,全面清晰定义所需解决的业务问题;第二步,调研企业内外部的源数据,确定本次分析活动所需的数据,数据的选择对模型技术的选择会产生重大影响;第三步,数据清洗,旨在发现并纠正数据文件中可识别的错误,包括检查数据的一致性、处理异常值和缺省值、剔除重复数据等;第四步,数据转换,完成更多的、更深入的数据处理;第五步,建模,应用多种模型技术,基于处理和转换后的数据,按业务需求完成各种不同类型的预测与细分;第六步,模型结果解释和评估,建模完成后将模型交给业务专家进行解释和评价,通过模型结果,发现细微的、有趣的业务规则、购买模式或重要的业务模式等。^[6]从已有的研究看,这样的步骤基本上是行业大数据应用的基本框架。对教育大数据来说,也可以按照这样的步骤来展开。

第四,洞察力、精准服务和基于数据决策的文化也是教育领域大数据应用的目标。各行业大数据应用的目标都是获得洞察力,提供精准服务和创造基于数据决策的文化。这三者的起点是洞察力,即从数据中获得有意义的信息,为决策奠定基础,决策的结果就是能够为客户提供精准服务,在市场上拥有竞争力。

大数据不仅仅是一种技术,它还是人们看问题、看现象的一种视角。文化意味着工作氛围、产品层次、资源配置、人际关系等,目前大数据显然对所有这些方面都产生着影响,特别是人们处理这些方面的问题时,需要大数据带来的洞察智识。因此,三者的关系是相辅相成、彼此融合的。这些对教育领域也特别重要。教育大数据的目的也是洞察教育的真谛,为学生提供精准的教育服务,改进教育决策,创设为每一位学生提供最好教育的文化。

第五,教育大数据研究必须考虑到教育问题的研究和解决,是跨领域的。事实上,行业领域的大数据应用是一个复杂的、经常跨行业的工程。对学习来说,学习始终是在正式与非正式、专业与非专业环境中进行的,因此对一个人的成长来说,他所生活的环境为他带来跨行、跨界学习的可能,学习的机会到处都是,学习数据的痕迹也到处都有。因此,我们面临的教育问题不单单是学校带来的,还包括很多方面,所以,教育数据的采集和分析也必须是多元的,才有可能真实、全面地反映学生的受教育状况和未来的发展潜能。

总体来看,教育大数据是面向教育改革的,不过“当下面临的变革并不是技术层面上的。这种改变影响着我们能够收集的数据类型以及我们对这些数据的挖掘方式,促使我们对学习、教学和获取知识的过程展开全新的理解”。^[7]因此,有必要对教育大数据的研究路径做深入的分析以发现其方向。

三、面向智慧教育的大数据研究路径

今天的教育已经走向智慧教育,它的核心目标是“培养具有高智能和创造力的人,(他们能够)利用适当的技术智慧地参与各种实践活动并不断地创造制品和价值,实现对学习环境、生活环境和工作环境灵巧机敏的适应、塑造和选择”^[8]。这样的人才培养,离不开智慧学习环境,而智慧学习环境的重要构成离不开最新技术的参与。

(一)智慧教育与教育大数据研究的关系

黄荣怀教授在他的《智慧教育的三重境界:从环境、模式到体制》一文中,归纳了智慧教育的基本特征,他指出智慧学习环境、新型教学模式和现代教育制度是智慧教育的三重境界。^[9]

在智慧学习环境中,除了传统学校管理系统能够提供学习者在校基本信息、学业成绩和成长的信息之外,一些最新的技术,如人工智能、虚拟现实、增强现实等也都逐渐进入教育领域,这些技术的其中一个收益就是它们一直都在产生大量的数据,这些数据能够

被搜集、分析和发现意义,它们也是我们理解和改善今天教育的最重要基础。

新型的教学模式也意味着更多地利用技术,正规的学校教学和学生自主选择的学习都会得到更多资源支持,特别是很多学校已经开展混合式学习,既有面对面的学习,也有在线学习,通常在线学习是面对面学习的有力补充,这些数据完全有可能被记录下来,上传到云端,可以帮助我们理解学生是怎样学习的,学习过程是怎样的,利用了哪些资源,有哪些问题,取得哪些成果,以便提供针对性的意见和建议。

现代教育制度的最基本特征是强调每个人有权利享受平等的教育,教育贯穿人的一生。学校教育阶段的学习只是漫长求知过程中的一小段经历。在今天这样一个时代,每个人自主选择继续教育的机会和资源很多。一个人的求学经历,学校会记录和存档,虽然现在的个人学习档案记录还不完备,但是从趋势上讲,由于终身教育成为很多人的选择,从教育制度上也应该有所保障,记录每个人的学习情况和学习成果。如上海市终身教育学分银行就可以对一个人继续教育(学历教育、职业培训和社区老年教育等)的学习进行学分认定,^[10]认定过程也就是数据搜集和整理的过程。

因此,今天我们谈智慧教育,它是一个从环境、模式到制度都能够得到数据支撑的完整过程,所以,在教育领域进行大数据研究,不只是发挥大数据对教育未来趋势的预测作用,还包括如何利用大数据提升教育智慧,这种智慧既来自个人、学校,也来自整个社会。

(二)面向智慧教育的大数据的价值

教育大数据研究应该是从大数据到大数据战略的追求过程。在教育领域,面向智慧教育的大数据应该是教学、学习、技术和文化互动的产物,数据、数据分析和数据应用方法论是教育大数据研究关注的核心。

我们在聚焦教育大数据研究路线图的设计时,需要明确一个问题:数据驱动的教育要实现的目标是什么?

教育领域的大数据应用的基本目标可以归结为通过一系列全面的数据,了解和预测学习者的学习行为并加以正向干预;如果从社会层面来理解,则是利用大数据提供的线索,更好地通过明智的决策优化资源配置,解决教育公平的问题。以这样的视角来考虑教育大数据的价值,可以归纳为拓宽对教育的认知、实现基于数据的教育决策、促进教育的进化和培育大数据文化等四个方面,如图1所示。

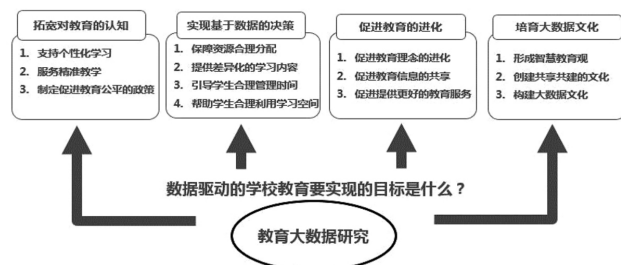


图1 教育大数据研究的价值

1. 拓宽对教育的认知

对学校而言,在智慧教育环境下,数据可以通过学校的管理平台、学习平台以各种形式被获取。这些数据可以被分析,用于评价学业成绩、发展状况,同样也可以帮助教师改进教学,帮助学生改进学习,帮助教育管理者改进管理。

教育大数据应用可以改善教育的三个方面:对学生而言,支持个性化学习;对教师而言,为精准教学提供服务;对教育管理者而言,为制定促进教育公平的政策提供依据。

首先,支持个性化学习。个性化学习的基础首先是个性化、具体化的学习内容。“大数据可以引领我们去理解知识的碎片,这样我们可以根据这些细小的碎片,一方面为学生提供量身定做的学习,另一方面可以发现每一个学生真正的问题在哪里……,它的效用会很大。”^[11]在大数据的帮助下,教师就有可能为学生提供他们最需要的个性化学习内容。除教师外,网络上各类课程层出不穷,学生通过网站和课程的反馈数据和评价数据,能够了解每一个平台及其网络课程的受欢迎程度,判断自己是否会喜欢某些平台的某些课程,然后选择符合自己学习 interests 的、高质量的课程。比如有一个专门的网站 coursetalky (<https://www.coursetalk.com/>),它根据学科、评分和费用等指标,请学生对现有的慕课平台提供的课程作出评价,基于一定算法,将评分最高的课程放在最上面,以此类推。选修课程的学生就可以发现到底哪些课是最受欢迎的,可以考虑如何选择学习。大数据在支持学生个性化学习方面,还有一点非常重要的是,它有可能为学生提供自适应学习的依据。自适应意味着计算机不断地评估学生的思维习惯,自动地为他提供量身定做的学习内容和材料。自适应学习软件为学习材料绘制相互联系的知识点地图,监控学生的每一项动作,记录学生看视频、阅读提示或学习材料、解决问题、做测试等每一次学习活动,然后根据一定的算法比较学生自己以及学生和其他人的学习状况,为学生提供适合自身水平的学习材料。^[12]

其次,服务精准教学。数据环境下,“精准教学过程产生的数据,其主要价值是服务于监测、评估学生的学习表现”^[13]。在计数与绘制阶段,教师将通过教学信息管理平台及其产生的数据获得非常准确的学生学习表现数据,他们将了解学生如何回答问题、回答的准确率怎样、用了哪些资源、哪些问题他们放过了、做了哪些研究、采取过哪些最有效的策略等,这些行为通过平台被即时、自动记录,教师就可以理解学习过程,在数据决策阶段通过分析他们的这些行为表现数据,为学生提供针对性的学习内容,甚至彼此契合的合作伙伴,帮助他们提高学习的效益,发现自己的潜能。教育大数据还能够使教师更全面地看待学习。传统的教学通过测试可以看到学生学业成绩如何,但是有了学生学习的总体数据,还可以帮助教师反思课程教学内容是否需要改进,从而为不同的学生订制、推送差异性的学习内容,在综合数据的引导下,改进学生学习和预测他们未来的发展方向。

最后,为制定促进教育公平的政策提供依据。其他行业的大数据应用非常强调“商业智能”,在教育管理领域的大数据应用将依赖“教育智能”,如果说商业智能是“将数据转化为信息,然后以此为基础制定决策,最终开始行动”^[14]。教育智能也是这样的要求,只不过基于数据的决策是面向获得学校管理的智识。由于今天正式学习与非正式学习都是人们的重要学习选择,每一个学生的学习选择都可以十分丰富。然而,由于各种差异,包括经济、思维、见识和能力等方面,会造成更大的数字鸿沟。而学校教育要在弥补数字鸿沟方面有所作为,就需要从管理决策入手,尽可能为所有学生提供使用大数据的机会,培养他们使用大数据的思维方式和能力。舍恩伯格和库克耶在《与大数据同行:学习和教育的未来》一书中提出:“如果教育中的大数据看起来在造就不公平的优势和促成新的不平等,那么公共教育政策就要跟进”,使之“不只属于最好的学校中最优秀的学生,它存在于任何地方”。^[7]

2. 实现基于数据的决策

教育大数据的应用应该指向预测学生的学习潜质,预防学生出现成长问题,为学生个性化学习提供支持,实现以学生为中心的教学,促进学生积极参与到学习中,并以较低的成本取得更好的教育成效。从本质上讲,基于数据的决策人可以是教育管理部门、学校、教师和学生自己等,因此决策的内容和层次会显著不同。这里主要说明学校基于数据的决策的一般应用。

在学校教育阶段,基于数据的决策目标主要包括四个方面:(1)为学生合理分配资源;(2)帮助学生选择合适的学习内容;(3)引导学生按照学习需要安排时间;(4)指导学生恰当利用不同的学习空间。要实现这些目标,可以通过不断追问一些问题获得洞察力来为决策奠定扎实的基础,如在资源、学习内容、时间和学习空间等方面,数据能够告诉我们什么?我们从中可以了解到什么?学生在网络留下的学习痕迹和学习结果的数据含义是什么?我们如何利用分析的数据指导我们的下一步行动?教师和学生能理解和接受我们的决策,并愿意和我们一起行动吗?这些问题的背后,揭示的是基于数据的决策的基本构成和关系,如图2所示。

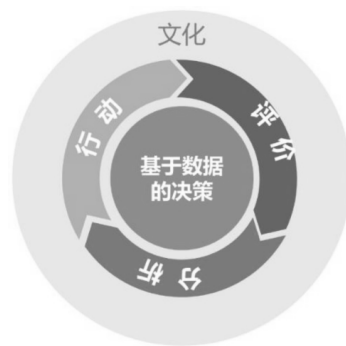


图2 基于数据的决策的基本构成和关系^[14]

当我们能够理解数据告诉了我们什么,并最后落实到基于数据结果的行动上,那么基于数据的决策就是有效的,而且从文化的层次来看,当师生也共同认可了这样的评价和行动原则、规范和方法之后,将会转化为自觉的行动。

还需要意识到,如果思考大数据对我们生活产生的影响,就会发现算法其实一直分析着我们的行为,不管是线上还是非线上。他们决定着我们会做什么,也常常操控着我们下一步做什么。因此,基于数据的决策不能完全依赖数据,需要在数据的基础上体现我们作为人的智慧,我们的情感、关系和需要等。

3. 促进教育的进化

目前已经有大量的学校管理系统、在线学习网站和系统、资源网站、学习APP能够记录和分享学习者学习的数据,我们如果能够通过利用工具和算法洞察数据给我们的启示,就有可能有更好的方法用于改进学校的教学。

大数据在促进教育进化方面可以发挥三种作用:

促进教育理念的变化。很多教育工作者都不太喜欢改变,而且不希望自己的教育权威被低估。然而大数据分析、学习分析带来了一些全新建议的学习方

法,如游戏化学习、利用 APP 的学习等,能够激发数字化原住民的学生的学习积极性,提高学习成效。这样有些人就很有可能认可新的以学生为中心的教学模式。学生自己也可以通过大数据分析自己的需求、社会发展要求、同龄人的选择等,发现自己的发展方向,改进学习策略与方法,成为乐于和有能力对自己学习负责的人。

促进教育信息的共享。今天的时代是一个共享的时代,产生大数据的各种平台围绕一人的可辨识身份识别编码,就能够汇聚有关一个人的综合信息,这些信息构成了他的数字化存在。从一个方面来说,大数据有可能带来对一个人隐私的侵犯,但就一个人作为学生的学习阶段,如果教师能够通过共享信息(在线学习平台、学校管理系统、社交网站等提供的个人信息)更多了解一个学生或者一个学生群体,教师就有可能采取更好的方法来帮助学生。

促进提供更好的教育服务。教育大数据的来源不仅仅是学校,还包括社会的各个领域,如社区、教育公司、博物馆等。过去学生参与社区活动,在教育公司的平台上学习,都是没有任何关联的。如今学校已经开始尝试通过电子学籍卡把这些信息聚合起来。当这些原先被割裂的信息整合起来之后,我们将有机会更加了解学生的需求、困难、收获等,以便提供更高效、更有符合个人需求和社会发展的教育服务。

4. 培育大数据文化

智慧教育,实际上是数据驱动的智慧社会的一部分。“我们究竟是怎样的一个人不是通过我们声称自己做了什么,而是通过我们去过的地点以及我们购买的物品等痕迹更为准确决定的”。^[15]这也意味着,我们的学习和学习的结果也是可以随时被记录、分享和评价的,构成了整个社会学习生态的一部分,而且个人的学习数据和其他活动数据一起将反映我们是怎样的人。这也使我们“处于一个证据驱动、数据驱动和分析性的文化”。^[16]因此我们的思考不能仅仅局限于学校教育系统或者技术平台的优势,而应从社会生态系统来审视智慧教育的可能和潜能。教育可以充分利用整个社会的多元化的资源来设计与实施。

教育大数据能够告诉我们这个世界的教育场景正在发生着什么,学生、教师、学校领导者以及关心教育的人对教育有什么样的看法,他们又是如何做的,有什么收获,有什么困难,需要什么资源支持等。需要看到的是,尽管大数据的作用不容小觑,但是通过大数据要聚焦的还是人的需求和发展。因此当我们强调进行大数据研究时,需要的还是从大数据中获得洞察

的智识,并且能够把这种智识转化为对教育设计与实施的基础,为培养今天的学习者创造智慧学习环境。同时,我们设计的教育系统通过多元化的数字工具和平台能够获取、分析和应用大数据,很多对大数据感兴趣的学习者也会逐步成长起来,利用大数据为教育服务。在这样的氛围中,大数据文化成为人类灿烂文化的重要组成部分。

(三)教育大数据应用研究路线图

如今,当我们拥有海量的教育大数据时,它的价值不在于大,而在于我们能够通过大数据分析模型和工具进行自动化的加工处理、挖掘、意义分析等,从中能够获得需要的教育智慧,提取我们所需要的教育启示。因此,我们需要依照教育大数据研究所涉及的任务和希望实现的目标对其进行规划,路线图如图3所示。



图3 教育大数据应用研究路线图

和企业一样,大数据是为解决问题以寻求更好发展而存在的,因此做教育大数据研究,第一步是理解智慧教育环境下,技术的参与要帮助解决教育领域的什么问题。教育领域的问题千千万万,对于很多问题虽然我们习以为常地相信某种结论,但在今天这样一个时代,需要数据来剖析其根本原因,比如成绩中等的学生容易受到教师忽视,数学成绩好的学生更聪明,这样的问题就比较适合用大量的数据样本来验证是否是真实的。

第二步需要思考这些问题涉及的因素是什么,可以从哪里获得相关原始大数据。做大数据研究的重点是数据的量要够大,数据具有一定的多元性,能够全面、真实、准确反映教育场景或教育场景中的人的全貌。

第三步,当厘清了各个要素和要素之间的关系,以及可以获取哪些数据来支持研究之后,需要建立分析模型。好的分析模型需要体现六大特征:(1)与要解决的教育问题有相关性;(2)具有统计学意义和预测力;(3)模型结果有可解释性、合理性;(4)能够兼顾运行效率,即在难度和工作量方面能够接受,特别是实

时在线评分模型,这一点尤为重要;(5)全流程运营的经济成本可以接受(事先做一个全面的成本效益评估很重要);(6)遵守本地和国际同行的相关法律法规。^[6]

第四步就进入教育大数据分析阶段,这是教育大数据研究的核心阶段。其过程涉及描述性分析,了解正在发生什么,潜在的趋势和发生的原因;预测性分析,确定未来可能发生什么;规范性分析,认识正在发生什么,对未来作出可能预测,同时制定决策以获得最大收益。^[4]其实实施大致需要从三方面着手:(1)分析条件:首先要引入大数据分析工具,因为大数据分析有很多重要环节,所以工具的选择非常重要,主要包括数据搜集工具(如 Import.io)、数据存储和管理工具(如 Hadoop, Spark)、数据清洗工具(如 OpenRefine, DataCleaner)、数据挖掘工具(如 RapidMiner, IBM SPSS Modeler)、数据分析工具(如 Qubole, BigML)、数据可视化(如 Tableau, Silk)、数据整合(如 Blockspring, Pentaho)、数据语言(如 R, Python)。^[17]对教育领域来说,还需要深层次地理解 xAPI,因为它提供了一个以学习者为中心的采集学习行为的数据模型。这些工具的应用需要很专业的知识和技能,所以教育领域有必要引入大数据科学家来参与大数据分析工作。当然,教育领域的专家必须共同工作,因为对教育的洞察需要专业知识。(2)大数据的分析涉及很多环节,包括数据采集、抽样和预处理,预测分析,描述性分析,生存分析,社交网络分析,等等,过程包括确定分析的要素,依据一定的算法对数据进行组合、关联、聚合,通过相互参照,发现数据的相关性。(3)分析结果:在相关性的基础上,利用数据可视化工具,以图像直观展示相关要素的关系,推演出一般的结论,然后从中获得启示。

以支持个性化学习和精准教学为例,美国的 Illuminate Education 公司的 DnA(Data and Assessment)系统,就是一个非常自动化的大数据分析系统,可以帮助学生自主学习和教师精准教学。目前,美国 45 个州 14000 多所学校的 500 万学生都在使用这个系统。系统的功能主要包括评价、分析和行动三大部分。评价指的是教师可以布置题库中与教学标准一致的题目,提供给学生练习,学生也可以自主学习;分析意味着教师可以使用含有大量明晰图表的、常规的或个性化的成绩报告来了解学生的学习走势,这些图形都是在算法的支持下自动整合进行分析的,为教学提供依据;行动是指教师可以根据学生报告的结果,对未达标的学生进行分组,然后可以为他们提供针对性教学,以便改进他们的学习成效。该平台能够跟踪学生的学习进展和干预效果,并提供汇总报告。这样学

生能够掌握自己的进展,教师一方面可以直观看到自己的教学效果,另一方面可以考虑新的教学干预。Illuminate Education 数据平台可以反映一个人和一个群体在学校学习的完整数据,学生、教师和学校的教育管理者都可以全面跟踪了解学生的学习状况和教学的有效性,作出基于数据的决策,支持学生的个性化学习。实际上,它可以做到从大数据中得到有意义的小信号,给学生、教师切实的帮助。该平台提供的许多案例都说明了其价值。如图 4 所示。

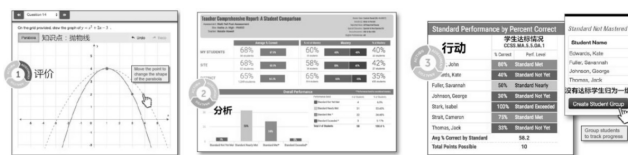


图4 Illuminate Education 数据平台示例^[18]

四、结 论

教育大数据也意味着我们有机会看到教育的大图景。通过它,我们可以了解教育的过去、现状和未来趋势,发现问题和解决问题。如前所述,教育大数据研究是考验人类智慧的一种研究,它是帮助我们理解教育相关的各个要素的重要途径。我们要做的是利用大数据研究展示人类智慧,寻求高水平地解决各种各样的教育问题的途径,推动教育朝着有益于学生成长和社会发展的方向转型。

其实,教师实际上是一直在使用数据,但是以往的数据来自他们自身的观察和测试,而大数据可以将这些放大很多倍,它们是成千上万的个体互动的结果,包括成千上万的学生每天正式学习和非正式学习的表现。这将消除正式学习与非正式学习原先的壁垒,丰富的学习机会直接能够互相补充,为学生提供完整的数字化的学习体验。因此,大数据有机会让好的教师成为更好的教师,提高教学质量和教学成效;让每一个学生清楚地知道,学习是可以以多种方式进行的,每一种都能够在一定程度上提高自己的学习成效,因此,要学会选择、学会担当,自己为自己的学习负责。

和所有领域的大数据一样,教育大数据也是从一个人(一个学习者)开始,然后经过数据的汇总、分析、得出结论再回馈给个人。我们可以视大数据为一种机会,一种个人发现自己和社会发生广泛联系的机会。在大数据的回馈下,学习者可以了解自己、了解社会、了解自己在今天这样一个时代何去何从。让每一个人成为最好的自己,这是我们教育的理想。教育大数据的确可以有所作为。

[参考文献]

- [1] 肖玉敏. 大数据来了,教育如何站位[N]. 中国教育报,2015-03-23.
- [2] JACKY AKOKA I C N L. Research on big data – a systematic mapping study[J].Computer standards & interfaces,2017.
- [3] RANKIN J G, JOHNSON M, DENNIS R. Research on implementing big data: technology, people, & processes[EB/OL]. [2017-01-31]. <https://eric.ed.gov/?id=ED555544>.
- [4] 沙尔达,德伦,特班,等. 商务智能:数据分析的管理视角[M]. 赵卫东,译. 机械工业出版社,2015.
- [5] KAYYALI B, KNOTT D, VAN KUIKEN S. The big-data revolution in US health care: accelerating value and innovation[EB/OL]. [2017-08-08]. <http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care>.
- [6] BAESENS B.大数据分析数据科学应用场景与实践精髓[M].柯晓燕,张纪元,译. 北京:人民邮电出版社,2016.
- [7] 维克托·迈尔-舍恩伯格,肯尼思·库克耶.与大数据同行:学习和教育的未来[M].赵中建,张燕南,译. 上海:华东师范大学出版社,2015.
- [8] 祝智庭,贺斌.智慧教育:教育信息化的新境界[J].电化教育研究,2012(12):5-13.
- [9] 黄荣怀.智慧教育的三重境界:从环境、模式到体制[J].现代远程教育研究,2014(6):3-11.
- [10] 上海市终身教育学分银行. 上海市终身教育学分银行[EB/OL].[2017-08-12]. <http://www.shcb.net.cn/>.
- [11] LEVINSON M. Personalized learning, big data and schools [EB/OL].[2017-08-13]. <https://www.edutopia.org/blog/personalized-learning-big-data-schools-matt-levinson>.
- [12] FLETCHER S. How big data is taking teachers out of the lecturing business [EB/OL].[2017-08-12]. <https://www.scientificamerican.com/article/how-big-data-taking-teachers-out-lecturing-business/>.
- [13] 付达杰,唐琳. 基于大数据的精准教学模式探究[J]. 现代教育技术,2017(7):12-18.
- [14] ADATO M. Setting up a data-driven culture [EB/OL].[2017-08-09]. <https://www.illuminateed.com/blog/2017/05/setting-up-a-data-driven-culture/>.
- [15] 彭特兰. 智慧社会:大数据与社会物理学[M].汪小帆,汪容,译.杭州:浙江人民出版社,2015.
- [16] STOUT K. A culture of evidence[EB/OL]. [2017-08-14]. https://www.sas.com/en_us/insights/articles/analytics/a-culture-of-evidence.html.
- [17] Import.io. All the best big data tools and how to use them [EB/OL].[2017-08-08]. <https://www.import.io/post/all-the-best-big-data-tools-and-how-to-use-them/>.
- [18] Illuminate Education. The Education Intelligence Platform[EB/OL].[2017-08-18]. <https://www.illuminateed.com>.

Research and Practice for Leveraging Big Data in Smart Education: Value and Path

XIAO Yumin, MENG Bingwen, TANG Tingting, WU Yonghe, XIE Rong
(Faculty of Education, East China Normal University, Shanghai 200062)

[Abstract] Big data is a kind of existence for more and more organizations and individuals trying to understand and take advantage of their strengths. Big data in education, just like what is happening in other fields, will help us to deepen our understanding of learners' growth, make data-based predictions for future possibilities of individuals or groups, and ultimately promote insights into human nature to provide better education to meet learners' needs. This paper explores the relationship between smart education and educational research on big data, and clarifies that the value of big data research in education is reflected in four aspects: broadening an understanding of education, realizing data-driven educational decision-making, facilitating the evolution of education, and cultivating big data-based culture. As a result, a roadmap of research on big data in education is designed and discussed.

[Keywords] Big Data in Education; Smart Education; Data-driven Decision-making; Roadmap