

解读教育数据智慧

祝智庭¹ 彭红超² 雷云鹤³

- (1. 华东师范大学 开放教育学院, 上海 200062;
2. 华东师范大学 教育学部教育信息技术学系, 上海 200062;
3. 上海市普陀区现代教育信息技术中心, 上海 200062)

[摘要] 大数据受到社会的普遍关注。教育领域的师生群体大,教育数据的潜在价值亦非常大。大数据的局限性日益显现,随后出现的小数据、全数据、快数据、行数据、关数据、巧数据等概念开始引起数据研究者的关注。本文从解读教育数据智慧的角度出发,理清了上述几个概念之间的关系,认为大数据与小数据组成了集数据密集与智慧密集于一体的全数据;快数据、行数据、关数据、巧数据是大数据的四种新形态,它们从不同的角度突显大数据的潜在价值。本文期望通过对数据智慧的解读,能够为教育数据的挖掘和应用提供思路。

[关键词] 数据智慧;大数据;小数据;全数据;快数据;行数据;关数据;巧数据;暗数据

[中图分类号] G40-057

[文献标识码] A

[文章编号] 1007-2179(2017)05-0021-09

大数据(Big Data)已经在21世纪掀起一场惊涛骇浪。根据国际数据资讯公司(IDC)的分析,世界上的信息正以每两年翻倍的惊人速度在增长。了解大数据、如何利用巨量资料,成了许多国家关心的重点议题。2012年3月29日,美国发布的《大数据研究和创新发展创新计划》提出28个研究项目,试图通过提高从大型复杂的数字数据集中提取知识和观点的能力,以加快科学与工程的步伐,加强国家安全,改变教学研究。欧盟制定了《数据价值链战略计划》,以实现数据的最大价值,重点是通过数据为中心的连贯性欧盟生态体系,让数据价值链的不同阶段产生价值。英国制定的《英国数据能力发展战略规划》,通过系统性研究,明确定义了数据能力(包

括人力资本、基础设施、软件和研发能力,数据资产)、如何提高数据能力,并提出了相应举措。2013年6月,日本公布的《创建最尖端IT国家宣言》报告,包括了开放数据、数据流通和创新应用等,在大数据与传统行业(如能源、交通、医疗、农业等)结合方面制定了明确计划。2013年,韩国提出建设开放大数据中心,对企业、大学和普通公民开放,支持利用大数据技术解决业务或研究问题。

《中共中央关于制定国民经济和社会发展第十三个五年规划的建议》提出:“实施国家大数据战略,推进数据资源开放共享”,将大数据战略提升到国家政策层面。国内IT巨头BAT(百度、阿里、腾讯)提供的大数据相关产品(百度大数据平台、阿里

[收稿日期] 2017-07-24

[修回日期] 2017-07-29

[DOI编码] 10.13966/j.cnki.kfjyyj.2017.05.003

[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”(BCA140051)。

[作者简介] 祝智庭,博士,教授,博士生导师,华东师范大学开放教育学院,研究方向:教育信息化系统架构与技术标准、信息化促进教学变革与创新、技术使能的智慧教育、面向信息化的教师能力发展、技术文化等(ztzh@dec.ecnu.edu.cn);彭红超,华东师范大学博士研究生,华东师范大学教育学部教育信息技术学系,研究方向:智慧学习生态、精准教学(hongchao5d@qq.com);雷云鹤,硕士,中学一级教师,上海市普陀区现代教育技术中心,研究方向:智慧教育与精准教学、基础教育信息化课堂改革(leiyunhe@126.com)。

云平台、腾讯大数据分析平台),可以帮助搭建大数据应用及平台。2017年6月29日,上海交通大学、中国联通小沃科技、中国联通研究院、北京广视通达共同筹建了“大数据联合实验室”,以提升大数据核心技术能力、深化大数据行业应用、推进产学研的标准化与规范化(黄辛,2017)。

大数据引起了各行业的普遍关注,正发挥着越来越重要的作用。教育是大数据实现潜在价值的重点领域之一,正以势不可挡之势推动着教育的变革。从我国印发的《促进大数据发展行动纲要》(2015年8月)可知,大数据已成为国家重要的基础性战略资源,引领着新一轮科技创新,其中专栏4明确提出了关于“教育文化大数据”建设的指示(中华人民共和国国务院,2015)。

一、从大数据到全数据的衍变

对大数据的大肆炒作,致使业界出现盲目的迷恋与跟风、一味地追逐数据规模的现象,忘记了自己是否有能力(或是否有能力承受发展大数据的成本)或有必要这么做。2013年12月,《哈佛商业评论》指出,许多企业不能管理与分析大数据,无法从中得出新见解,它们或许压根儿就不需要大数据(Ross et al., 2013)!其实,大数据仅仅是数据科学的一角,除此之外,还有小数据(Small Data)和全数据(All Data),对应到教育领域,即为教育大数据、教育小数据和教育全数据(见图1)。

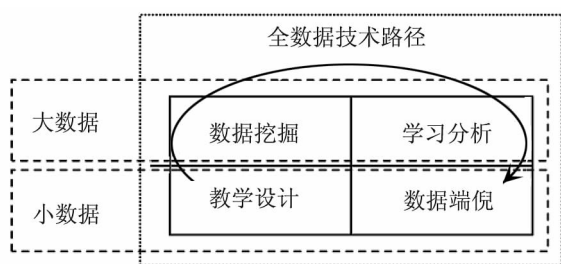


图1 全视角的教育数据

(一) 挖掘教育大数据的价值

在《智能时代:大数据与智能革命重新定义未来》(吴军,2016a)中,大数据和机器智能领域资深专家吴军博士将2005年定义为大数据元年,这是因为,2005年美国国家标准与技术研究所对全世界多种机器翻译系统的评测中,Google在机器翻译专家

弗朗兹·奥科(Franz Och)博士的主持下,使用上万亿于通常的数据量,训练出了一个六元模型(当时大部分研究团队只能训练三元模型),这一事件引起了研究者对数据量的关注。2008年,《自然》杂志出版大数据专刊(Nature, 2008),对大数据的相关技术和挑战进行了研讨,并提出了Big Data概念。此后,大数据开始受到关注。

大数据因其巨量、高速、多样而需要用特殊的技术和分析方法(传统的数据处理程序不足以处理(Snijders et al., 2012),才能将其转化为有价值(如增强决策、洞悉发现、过程优化(Gartner, 2012))的信息资产。《大数据的人性面孔》(Smolan & Erwiit, 2012)给出了一个哲学定义:大数据是帮助地球建构神经系统的过程,在这一系统中,我们(人类)不过是其中一种感测器。大数据具有6V属性(Demchenko & Gruengard, 2014):巨量(Volume)、高速(Velocity)、多样(Variety)、稀值(Value,价值密度低)、真实(Veracity)、迅变(Variability)。当然,并非同时具有六大属性的数据才是大数据,特别是教育大数据。从大数据中挖掘出的信息,通常是传统统计方法和常规数据工具无法获得的。比如,从百度知道获得的数据可以得出各地人们的饮食习惯。大数据事实上将智能问题转变为数据处理问题。比如,在人机围棋大战中,Alpha Go通过大量的棋谱学习,最终战胜了人类。大数据应用的一个典型案例是交通信息实时路况(如Google的城市智能交通管理):在大数据出现前,交通信息的指挥始终处于延迟状态,大数据出现后,实时路况播报成为新常态。

大数据作为一种信息资产或资源,具有超乎物质资源的特性:非竞争(non-rival),可以被用户同时使用;可再生(renewable);可以被重复、循环使用并按指数持续增长;多用途(multi-purpose),同一大数据可被诸多行为主体做不同解读,实现不同的用途。这些特征在促进教育科学化的同时,也使安全与隐私问题变成巨大的隐患。

处理或分析大数据的相关技术称为大数据技术,它旨在处理巨量、高速、多样的数据(集/资产),以提取潜在的数据价值,确保原始信息和已有信息的高精确性。这要求具有成本效益、创新式的数据和信息处理(分析)技术,以提高洞察、决策和过程控制。所有这些要求新的数据模型(支持数据生命

周期中的所有数据状态和阶段)和新的基础设施和工具(的支持),以从多样的来源(包括感知器网)获取(和处理)数据,并以多种形式传递数据给不同的数据(信息)消费者和设备。大数据技术的教育应用为教育数据挖掘与学习分析技术(见图1)。前者侧重于大数据潜在价值的萃取、学生相关模式的识别等,如回答何种主题编列对特定学生最有成效;哪些行为与更有效的学习或更高成绩有关;哪些行为表明学生参与了学习或对学习满意;学习环境的哪些特性导致了更有效的学习等。后者在教育数据挖掘提炼出的价值以及识别出的模式的基础上,为教育决策提供精准、适性的服务(祝智庭,孙妍妍等,2017)。

(二) 寻找教学的小数据痛点

大数据的6V属性,使其富有巨大潜在价值的同时,也极难处理或解析。许多公司投入大量资金到复杂的数据分析工具后,才发现他们没有能力解释大数据,从中得出新见解,也没有能力将它变成竞争优势。因此,做大数据,最好的策略是从小数据做起。特别是在教育领域,学会如何利用小数据,是遨游日益复杂的数据世界所需的基本技能。

小数据是具有小体量(Volume)和特定属性(Kavis, 2015)的数据,这使它易于理解、便于访问、可操作性强,从而可以对日常事宜做出及时、有意义的反馈(Rouse, 2014)。体量方面,它可以非常小,小到仅仅是一个数值。属性方面,它可以具体到某一微不足道的行为观察。

小数据其实也有大见解。马丁·林斯特龙(Lindstrom, 2017)2004年负责为陷入亏本的乐高公司做整体品牌战略咨询(此前,乐高公司根据大数据研究分析得到结论“未来几代人会对乐高积木失去兴趣,因为数字原住民没有时间和耐心玩乐高”,因此将主要精力转向主题公园、视频游戏等新市场),他们见到德国中型城市的一个11岁男孩(不仅是乐高迷,还是狂热的滑板爱好者),问他最钟爱的东西,得到的答案是:那双破旧的阿根达斯运动鞋是他的战利品、金牌和杰作。乐高公司由此意识到,孩子们要想在同龄人中获得社会存在感,就要具备一种高超的技能。无论这种技能是什么,只要值得花心思、花精力去做。此后,乐高重新回归核心产品——积木,并于2014年成为全球最大的玩具生

产商。显然,大数据分析得到的结论让乐高陷入危机,而“德国11岁男孩的旧运动鞋”这一小数据却给乐高公司的定位带来极大转变。虽然,小数据往往隐藏在生活小细节中,看似杂乱无章、毫无导向、微不足道,但实际上一样能够具有突破性启发和颠覆性影响的价值。在大数据的热潮中,越来越多的企业将注意力放在数据规模上,忽视了最初的真目的:解决关键的业务问题或挑战。

小数据不同于大数据的地方在于:数据的体量、数据的种类和格式、数据的处理速度和数据的复杂度。大数据属于监测数据,是全样本的、监测记录的、客观的、过程的、连续的,而小数据属于调查数据,是抽样的、样本反馈的、主观的、结果的、断点的。这些可以作为区别小数据与大数据的原则。英国医院俱乐部举办的2016小数据论坛提出了五点洞见,很有参考价值(Quinn, 2013):第一,让数据有价值,是关乎适当的各成分相融合的问题;第二,大数据是石油和土壤,但如何找到对数据内在设计的深度理解至关重要;第三,撬开坚果需要的不是大锤,而是胡桃钳,良好的洞悉不一定来自整理和处理越来越多的数据,而是对已得数据更加灵活的解析;第四,更好地了解当前的环境,能够创造更美好的未来;第五,算法很重要,但是人类要素是无法复制和替代的,因为小数据是具有情感温度的(Sarkar, 2016),算法在这方面无计可施。

小数据的教育应用可归于两点:教学设计和数据端倪(见图1)。本专业素有教育技术是“姓教”还是“姓技”之争。其实,相比这二者,教学设计更为重要,这直接关乎技术是否能够助力、增能教育,带来更富有成效的效果。与大数据不同,小数据更注重用户、更强调目标,这使教师可以事先开展数据驱动的教学设计,制定最佳的实施方案,可以事前确定哪些数据的获取与分析是必要的。对于教学得到的小数据,可以在统计分析的基础上,采用数据启发的专家研究法(即数据端倪法,系本文作者首倡),来获得突破性的关键见解。

可穿戴设备虽然已有走进课堂之势,但很多数据依然无法实现数字化,机器的智能目前也无法代替人类的智慧,这就为小数据提供了切入点。比如,班主任及学科教师对班级每个学生都有比较深入的认识,基本了解班级同学的听课状态、作业情况、思

维方式、学习习惯、心理承受能力等。这些信息是教师与学生接触中获得的,但往往不能完全反映到学习成绩中。这些数据对于班级管理和学科教学有重要作用,如某同学成绩大幅下滑,班主任发现他上课心不在焉、抑郁寡欢,由此可通过观察、交谈、随访等找出原因,制定适切的策略,帮助该生恢复学习的热情与激情,进而实现提升成绩的目标。而这些观察、交谈、随访得来的大数据通常难以通过大数据获取。

(三) 构建全数据的教学策略

通过上述对大数据与小数据的解析可知,它们均有各自的优势与不足。因此,将大数据与小数据相结合,实现优势互补,就形成了接近“全数据(All Data)”的解析。大数据的6V属性,注定其处理只能由机器来完成,并且常规的数据存储(如关系型数据库)与分析(如统计分析)技术亦无法达到大数据的要求。而小数据的情感温度特性往往需要线下、通过亲身观察、交谈、随访等获得,因此小数据更偏向于由人来完成,借助专家的智慧来洞悉背后隐藏的价值。从这个角度讲,全数据体现的是人机协同的数据智慧:人的智慧密集与机器的数据密集的协同智慧。

大数据重在“相关关系”的解析,它放弃了对“因果关系”的追求,即大数据解决“是什么”问题,而不解决“为什么”问题,这大大降低了概率预测的成本。最近研究也提出了一种担忧:大数据可能落入所谓的“醉汉路灯下找钥匙(Drunkard's Search)”或“路灯效应(Streetlight Effect)”。有这么一个故事:醉汉在路灯下不停地转来转去找东西,路人问他丢了什么,他说家门钥匙丢了(感兴趣的问题)。路人帮他一起找,结果转了几遍都没找到。路人就问,你在哪里丢的钥匙?醉汉说,我出了家门(数据样本1)钥匙就丢了。路人大怒:那你到这里来找什么?醉汉振振有辞:因为只有这里(数据样本2)有光线啊!大数据的“路灯效应”表明,数据样本或规模的庞大也未必能够保证产生我们最感兴趣问题的见解,如果没有适当的样本,往往得到不想要的结果,甚至会产生错误的见解。因此,超越大数据走向全数据分析是避免“路灯效应”的必然要求。

另外,数据体量越大,噪声的比重会越大、信号的比重会越小,因此越容易将噪声(比如,随机性相关)误认为是信号,这种没有因果关系的相关性很

容易导致事实的假象(Walker, 2013)。小数据重在“因果关系”的解析(Lindstrom, 2016),这种因果关系与大数据相关关系的概率预测不同,它侧重于对过去的解读,即用过去的数据说明过去。这种通过因果关系得到的见解只是一种知识或经验,而我们希望利用已有的数据,解决未来会遇到的问题或挑战,因此,这种知识或经验需要提升至智慧(智慧是一种解决前所未遇的问题的能力(祝智庭 & 彭红超, 2017))层面。从这个角度讲,小数据需要分析者自我突破,将知识与灵感融合,方能发现价值。

综上,大数据与小数据结合形成的全数据,是数据密集(机器)与智慧密集(人类)的结合,是相关关系与因果关系的结合,是未来预测与过去解读结合的产物。在教育领域(见图2),大数据体现出机器智能的优势,小数据侧重于运用专家智能。大数据与小数据的有效结合,为人机协同的智慧教育提供了可能,也为我们引出一一种发掘教育智慧的可行路径,即:教学设计(小数据)→数据挖掘(大数据)→学习分析(大数据)→数据端倪(小数据)(见图1)。其中,教学设计要明确需要解决的关键教育教学问题是什么;制定相应的教育教学策略与方案;实施策略与方案需要哪些方面的数据支撑(核心数据面);以核心数据面为中心,将周边的相关数据面考虑进来并选取相应的数据源(小数据跃升为大数据)。数据挖掘通过业务理解、数据理解、数据预处理、建模、模型评估、模型部署为学习分析提供高质量的数据模式。学习分析基于数据挖掘提供的数据模式,对学习进行描述、诊断、预测和处方,以保证教育教学的总体发展趋势始终指向预期目标(得到较为宏观、处方的适性策略),解析出数据间的相关关系;基于相关关系进行降维,并按照拟解决的关键问题与实际意义将大数据划分为多个小数据集(大数据分解为小数据)。在数据端倪中,以学生个体或几个学生为对象,通过观察、交谈、随访等方法进行细微的关注;结合学习分析分解形成的小数据和本阶段得到的细微数据,形成新的数据集;教师依据新数据集,开展数据启发的专家智慧密集研究,从而得出深入的见解(获得操作性策略)。

大数据和小数据所蕴含的教育智慧逐步显现,随即出现的快数据、行数据、巧数据、关数据、暗数据等也同样不容忽视(见图2)。图2的横轴左半轴代

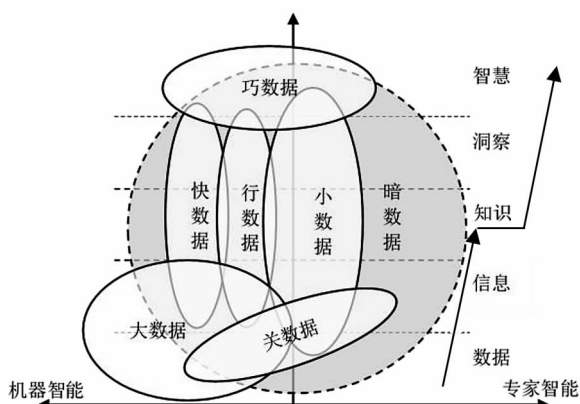


图2 数据智慧分布

表机器智能,右半轴代表专家智能。大数据是从数据中发掘价值,更多地体现为机器智能,小数据主要依赖专家的独特专业视角。快数据和行数据体现出发挥机器智能的方式,暗数据(Dark Data)、关数据和巧数据是机器智能和专家智能有效结合的智慧策略。纵轴代表教育数据形态在数据、信息、知识、洞察、智慧等不同阶段所处的位置。其中,暗数据的边界几乎无限,可能延伸至所有区域和不同阶段。

二、教育数据涌现四种新形态

高德纳咨询公司(Gartner)提出的著名的技术成熟曲线(the Hype Cycle,炒作周期)分为五个阶段:创新蒙动期(Innovation Trigger)、预期膨胀期(Peak of Inflated Expectations)、幻灭低估期(Trough of Disillusionment)、开悟复苏期(Slope of Enlightenment)、生产高峰期(Plateau of Productivity)。2011年,大数据进入成熟曲线的创新蒙动期,2012年进入预期膨胀期并于2013年达到最高峰值,2014年跌落幻灭低估期。此后,大数据离奇地消失在成熟曲线中。很多人由此得出结论:大数据已死!

无论大数据有没有死亡,可以肯定的是,大数据没有消失,而是呈现出四种新形态(Tonyshan, 2015):快数据(Fast Data)、行数据(Actionable Data)、关数据(Relevant Date)、巧数据(Smart Data)。这或许是人们在探究大数据的应用、萃取大数据见解并将其转化为行动过程中形成的产物。作为教育全数据的重要组成部分,教育大数据的这四种新形态也值得探讨。

为了便于理解各类教育数据,本文借助“小木

屋隐喻”(见图3)来说明:快数据可以看作是阳台,其价值易受外部因素的影响,且最易晒出教学的反馈情况;行数据可以看作是屋梁,对教学智慧具有直接支撑作用;关数据可以看作是骨架,支撑、联通各类教育数据;巧数据可以看作是墙体,是数据特征、模式的集中体现。而小数据可以看作是门窗,是数据驱动的教学策略的进出口和通气口。无论大数据还是小数据,均含有暗数据,它可以看作是地基,虽然处于黑暗状态,但价值潜能巨大。收集到的大数据、小数据,在未(能)萃取出所需的价值前,均可认为处于暗数据状态。在大数据、小数据的支撑下,教育智慧可以得到更好的体现,因此,可以将教学智慧看作是屋顶。

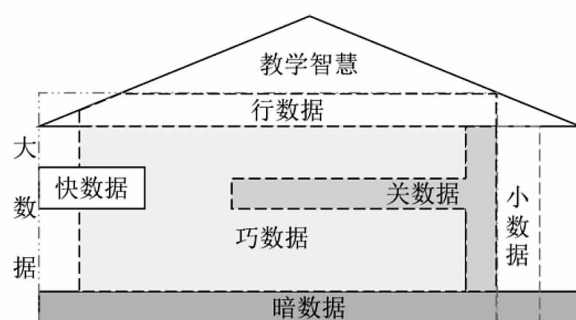


图3 教育数据的小木屋隐喻

“小木屋隐喻”初步展示了各类教育数据的功能,接下来进一步解析大数据的四种数据新形态。

(一) 快数据:教学反馈的加速器

快数据,是大数据中快速流动的、时间敏感的、需要快速洞悉、即时决策、立即付诸于行动的数据。快数据的流动速度等同于网速,使用快数据的最大挑战是时间(Duckworth, 2016)。因为,快数据是时间敏感的,自生成后,其价值在迅速流失,因此常常是几秒甚至几毫秒就要洞悉出见解,做出“做什么、如何做”的决策,并立即付诸于行动(比如监测流量、跟踪疫情、股票交易等)。当数据流入长期分析与存储引擎后,就失去了实时处理的机会。快数据的动态,可以采用“数据仪表盘”技术可视化呈现。处理快数据的技术(主要是流技术)需要能够实时处理高速度、大体量的数据,以使决策者能够及时了解新趋势和新变化,从而在正确的时间,访问正确的数据,进而采取正确的行动(Ashjian, 2015)。因此,快数据技术具有两大属性:即时性与精准性(Kumb-

la, 2016)。

快数据应用场景非常多。例如,公关公司需要实时了解客户对品牌的谈论,以便将负面言论扼杀在萌芽中,以避免其病毒式传播。公共卫生人员需要及时了解疾病的爆发,以便可以采取行动阻止蔓延。银行需要紧跟地缘政治和社会经济形势的瞬变,以便在全球宏观政策下做出最佳决策。物流公司需要实时了解道路状况(如拥堵),以便及时调整行车路线。这些场景中的数据,均需要快速精准处理,每分钟的流失,均可能使其无法控制或失去价值。

在教育领域中,实时反馈并不一定是必须的,但实时反馈有时能很好地促进学习的高效达成。例如,某些在线英语学习平台,能够根据学生的口语发音,即时判断学生发音是否标准,随时给学生提供反馈,这就解决了英语口语教学中反馈滞后的问题,为学生口语学习提供了更好的脚手架。实时反馈通过即时的自动化反馈,极大提高了学生的学习效率。

(二) 行数据:教学分析决策的利器

行数据,是可资(actionable)行动数据的简称,它是使用时可促使使用者采取系列行为的数据(McManus, 2016),如报告数据、分析数据、仪表盘中的数据。行数据强调数据转化为洞察力、行动的重要性(I-SCOOP, 2013)。它通常采用综合预测分析和假设分析等方法给予使用者建议或处方,以使其采取反馈措施。例如,社会化分析使企业能够提炼隐藏在社交数据和活动背后的意义和潜在价值,以便收集可资行动的见解。总之,行数据可以促使使用者采取优雅、有效的行动,从而“到达你想到地方”。

为了确保行数据可资行动,必须确保它是可用的,因此需要做到以下几点:1)数据的组织与呈现尽可能清晰、直观、易懂;2)数据应能够易访问、可重复使用;3)数据具有一致性(如一致的格式、一致的定义),且可互操作。为了确保行数据可用,必须确保它是可信的,因此需要做到以下几点:1)确保数据来源渠道是可信的;2)具有合理的数据收集维度指标,避免收集不相关的字段(这属于噪声),以能够刻画精准的用户画像或事态局势;3)采用适宜的数据收集方法;4)定期清洗数据,以保证数据的高质量;5)定期检验,以保证数据准确。

行数据在教育中尤为重要,利用数据来“资助”教与学的行为或活动的优化,是大数据价值的重要体现。教育领域的行数据主要指可支持和辅助教育教学决策的数据。例如,衢州书院中学的老师可以根据学生学习成绩、每个学生的详细数据,汇总分析班级学生的情况,并据此总结提炼需要改进的教学方法,帮助学生获得学业上的精进和提高。

(三) 关数据:教育信息化的黏合剂

关数据是表征数据或数据集间关系的信息,它可以将数据从一个事件或活动传递到下一个事件或活动(IBM, 2017)。关数据以现实意义为准则,即数据间的(显式、隐式)相关性均是有现实意义的。关数据通常有三种形态:1)相关分析得出的数据;2)实现数据松耦合的规范数据;3)链数据(Linked Data)。

大数据侧重于对“相关关系”的解读,解读出的数据即为关数据,这种形态的关数据可助力大数据实现高精度的概率预测,从而提供个性化学习服务。基于相关分析得来的关数据,原始样本要足够大,从而避免假性相关。大数据的相关性分析正应用于很多领域,如谷歌内容广告(AdSense for Content)服务利用大量数据统计,做出相关性分析,使广告和内容的搭配效果更佳;电影租赁和收看视频的网站与零食广告;咖啡评论和销售网站与信用卡和房贷广告等就是其中的案例(吴军, 2016b)。

随着大量的教育APP与平台涌入各级各类学校,多数学校均同时并行使用多个APP或平台(有的实验校使用了十多个平台与APP),不同APP与平台间的数据联通成了亟待解决的难题。而能够实现这些APP与平台之间数据松耦合的规范数据将是关数据新的发展方向。教育部教育信息化技术标准委员会已经启动了“人人通”标准的研制,其中一项重要工作即是上述规范数据的研制。对于链数据,是语义关联数据,较为简单的是使用诸如HTTP、RDF或URIs的Web技术进行语义关联。

(四) 巧数据:教育数据的显影液

在数据分析盛行的时代,数据只有被智慧地使用才能发挥价值,社会更需要的是巧数据分析。巧数据指有实际意义的、通过智能算法提取出的特征数据、模式数据(Lorentz, 2013)或有其他巨大价值的。从大数据的稀值属性讲,巧数据就是价值

密集(I-SCOOP, 2014)的那部分的大数据区块数据。如果没有额外的智能层,收集大量的统计数字几乎是无效益的。巧数据分析是结果导向的、投资回报率(ROI)为中心的、敏捷的、及时的、有意义的分析(Matillion, 2015)。它可以“按需分析”,并以可资行动的时间尺度提供有见解的、以业务为中心的方案。通过巧数据分析,大数据可快速转化为行数据。

在教育领域,通过智慧方式取得并用以辅助教学的数据即为巧数据。数据只有以智慧方式使用时才更有用。例如,浙江衢州市书院中学的学生,在期末考试后都可收到一份长达15页、“私人定制”的“学业诊断分析报告单”,对每位学生一学期的成绩进行大数据分析,描述各科目各知识点的掌握程度,并对个人能力做综合评价,就像一份全面的“体检诊断报告”。据此,学生可以发现自己的不足,提高学习的针对性和有效性;教师可以发现教学问题并及时改正。例如,班级成绩分析包含基础成绩表、成绩分段统计、三率统计、合格率统计、优秀排行榜、学科均衡度分析、题目得分详细、S-T分数统计分析、超均率统计分析、分数百分等级分析、高低分组分析、学生成绩单、低分率、优良率等,为学生的成绩报告单提供全面的数据基础。

另一个案例是上海卢湾一中心小学开发一款云手表,用于监测学生的血氧含量和心律。教师对学生的云手表数据进行实时监测,能够及时地针对不同的学生体质,调整身体锻炼方案,以免运动不当而受伤。

三、教育数据面临的挑战

(一) 如何照亮教育暗数据

教育领域涉及庞大的师生群体,包含正规教学体系的教师和学生、非正式学习群体(如网络课程的自学群体)等。近年来,教育信息化工作的广泛开展,由此积累了大量的基础数据和学习数据。这些数据基本处于“沉睡”状态,如何充分“唤醒”这些数据,发掘它们的价值,成为迫切需要解决的关键问题和挑战。

学校中持有的、但尚未被分析或利用其全部潜在价值的数据其实是一种暗数据。这种暗数据可以称之为I型暗数据:休眠数据(Sleep Dada)。除了

I型暗数据,还有另外两种暗数据:II型暗数据——渗漏数据(Leaking Data)、III型暗数据——隐秘数据(Hidden Data);渗漏数据指以意想不到的方式进行分析,并产生“发布者”没有预料到结果的数据。隐秘数据指无法直接看到的数据,但可以通过观察其对周围环境的影响来推断其存在的数据。(Jamie, 2017)

产生暗数据的原因很多,如:1)由于技术过时,数据变得无法访问;2)数据被遗忘,且从未被分享过;3)数据收集了,却无法检索得到;4)数据虽然存在也能检索,但不能带来实际效益的成本数据;5)能够收集但因成本或技术问题而不收集。大部分暗数据只被使用过一次,就被储存在逻辑混乱、数量庞大的其他社群中。实际上,一个信息系统产生的80%的文档都未被二次使用。

高德纳咨询公司把暗数据当做一种积极的“闹钟”,提醒企业应该改变未被利用的数据的现状。福布斯(Forbes)认为暗数据由过街老鼠转变为香饽饽的唯一办法,是将这些数据灵活运用在其他业务中。暗数据产生于不断增加的复杂的加工进程中,由此,这使之也具备规划未来产业形态的能力。

在教育中,学生的心理发展程度、认知发展程度、先前知识背景和结构、学习过程中的行为、习惯、动作和偏好等数据不易发现,或者暂且不能发掘出价值。另外,也存在一些目前技术手段无法采集的数据,或者尚未被关注的的数据。这都需要借助认知心理学的理论进行深度提炼。这些暗数据一旦发掘出价值,将对教师的教学指导起重要作用。因此,“如何照亮教育暗数据”成为需要解决的重大问题。

(二) 解密教育数据的智慧

前面讲到了大数据、小数据、全数据、快数据、行数据、关数据、巧数据、暗数据。无论何种数据,都是没有任何价值偏向的(即数据具有价值中立特性),都需要由数据提炼为信息、提炼为知识和智慧,从而更好地为教育教学服务,这个过程即是智慧教育中的数据智慧机理。然而,不同类型的数据对教学的价值各异(见图4)。从大数据到巧数据分别处于不同的数据价值层级,解决的问题分别对应于知几无(know-nothing)、知是何、知如何、知为何、知何为等不同阶段。从不同数据对教学决策的支持程度看,纵轴向上表示显在的程度高,对教学的支持

程度大,纵轴向下表示潜在的教学意义大。如何有效地发挥教育数据的教学价值,是智慧教育面临的极大挑战,也是我们团队一直努力攻克的课题。

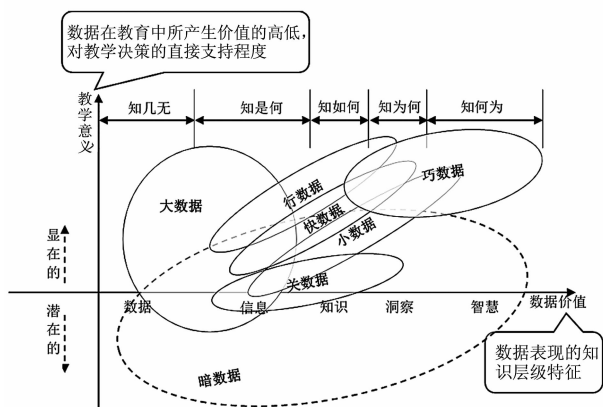


图4 教育数据价值分布

有研究者为了解决大数据面临的问题,将大数据方法的便捷性和小数据方法的精确性相融合,称之为全数据革命(all data revolution)(Yoshida, 2014)。教育数据智慧的获取,需依靠数据挖掘(本文指知识发现,包括数据清理、数据集成、数据变换、数据挖掘、模式评估、知识表示)和有效应用,这就要求教师增加一个新角色:数据科学家。具体来说,这就是要求教师适当采取编程和统计技术等手段(或者在编程和统计人员的帮助下),寻找全数据中蕴含的教育意义和契机,合理决策,改进教学。在教育教学、培训及学习中,挖掘大数据,关注小数据,照亮暗数据,整合快数据,转换行数据,挖掘关数据,提升巧数据,成为通过教育数据智慧的必经之路。因此,未来的研究主要关注以下几个问题:寻找并发掘特定教学场景下的教育数据模式;教育数据环境下教师角色的转型发展;教育数据背景下的(正式和非正式)学习方式。

[参考文献]

- [1] Ashjian, L. (2015). Machine intelligence + Human insight = Fast data [EB/OL]. [2017-07-13]. <https://www.tibco.com/blog/2015/03/27/how-analytics-facilitates-fast-data/>.
- [2] Demchenko, Y., & Gruengard, E. (2014). Big data course and learning model for online education (LMO) [EB/OL]. [2017-07-07]. <http://www.uazone.org/demch/presentations/edison2014wsh-big-data-edu-v01.pdf>.
- [3] Duckworth, D. (2016). What kind of industries will expand

their use of fast data in 2016? [EB/OL]. [2017-07-13]. <http://data-economy.com/2016/03/what-kind-of-industries-will-expand-their-use-of-fast-data-in-2016/>.

[4] Gartner (2012). Gartner says big data creates big jobs; 4.4 million IT jobs globally to support big data by 2015 [EB/OL]. [2017-07-13]. <http://www.gartner.com/newsroom/id/2207915>.

[5] 黄辛 (2017). “大数据联合实验室”在上海交大成立 [EB/OL]. [2017-07-12]. <http://news.sciencenet.cn/htmlnews/2017/7/381536.shtm>.

[6] IBM (2017). Relevant data [EB/OL]. [2017-07-13]. https://www.ibm.com/support/knowledgecenter/en/SSTFWV_5.1.0/com.ibm.itim.doc/cpt/cpt_ic_wkflo_data_relevant.html.

[7] I-SCOOP (2013). What makes data actionable data? [EB/OL]. [2017-07-13]. <https://www.i-scoop.eu/big-data-action-value-context/data-actionable/>.

[8] I-SCOOP (2014). From big data to smart data, processes and outcomes [EB/OL]. [2017-07-13]. <https://www.i-scoop.eu/big-data-action-value-context/from-big-data-to-smart-data-processes-and-outcomes/>.

[9] Jamie (2017). What is dark data? [EB/OL]. [2017-07-10]. <http://www.exploringdatascience.com/general/dark-data/>.

[10] Kavis, M. (2015). Forget big data - small data is driving the internet of things [EB/OL]. [2017-07-13]. <https://www.forbes.com/sites/mikekavis/2015/02/25/forget-big-data-small-data-is-driving-the-internet-of-things/#56aef7ad5d7e>.

[11] Kumbla, S. (2016). Fast data: Powering real-time big data [EB/OL]. [2017-07-13]. <http://www.cuelogic.com/blog/fast-data-big-data-in-real-time/>.

[12] Lindstrom, M. (2016). Why small data is the new big data [EB/OL]. [2017-07-13]. <http://knowledge.wharton.upenn.edu/article/small-data-new-big-data/>.

[13] Lindstrom, M. (2017). Small data: The tiny clues that uncover huge trends [M]. Manhattan: St. Martin's Press: 1-17.

[14] Lorentz, A. (2013). Big data, fast data, smart data [EB/OL]. [2017-07-13]. <https://www.wired.com/insights/2013/04/big-data-fast-data-smart-data/>.

[15] Matillion (2015). Goodbye data analytics, hello smart data analytics [EB/OL]. [2017-07-13]. <https://www.matillion.com/insights/goodbye-data-analytics-hello-smart-data-analytics/>.

[16] McManus, P. (2016). Smart data is actionable [EB/OL]. [2017-07-13]. <http://www.linkedin.com/pulse/smart-data-actionable-b2b-data-expert-intelligent-data>

[17] Nature (2008). Big Data [DB/EL]. [2017-08-22]. <http://www.nature.com/news/specials/bigdata/index.html>.

[18] Quinn, M. (2013). Forget big data —let's talk about all data [EB/OL]. [2017-07-13]. <https://venturebeat.com/2013/09/03/forget-big-data-lets-talk-about-all-data/>.

[19] Ross, J. W., Beath, C. M., & Quaadgras, A. (2013). You may not need big data after all [EB/OL]. [2017-07-13]. <https://>

hbr. org/2013/12/you-may-not-need-big-data-after-all.

[20] Rouse, M. (2014). Small data [EB/OL]. [2017-07-08].

<http://whatis.techtarget.com/definition/small-data>.

[21] Sarkar, C. (2016). "Small data, big impact!" - An interview with Martin Lindstrom [EB/OL]. [2017-07-13]. <http://www.marketingjournal.org/small-data-big-impact-an-interview-with-martin-lindstrom/>.

[22] Smolan, R. & Erwit, J. (2012). The human face of big data [EB/OL]. <https://www.engagespark.com/blog/big-data-the-potential-and-the-pit-falls/>.

[23] Snijders, C., Matzat, U., & Reips, U. D. (2012). "Big data": Big gaps of knowledge in the field of Internet Science [J]. *International Journal of Internet Science*, 7(1): 1-5.

[24] Tonyshan (2015). Big data is really dead [EB/OL]. [2017-07-09]. <https://www.smartdatacollective.com/big-data-really-dead/>.

[25] Walker, M. (2013). Small Data is beautiful [EB/OL]. [2017-07-13]. <http://www.datasciencecentral.com/profiles/blogs/small-data-is-beautiful>.

[26] 吴军(2016a). 智能时代:大数据与智能革命重新定义未来[M]. 北京:中信出版社:57.

[27] 吴军(2016b). 智能时代:大数据与智能革命重新定义未来[M]. 北京:中信出版社:132-134.

[28] Yoshida, N. (2014). Revolutionizing data collection; From "big data" to "all data" [EB/OL]. [2017-07-10]. <http://blogs.worldbank.org/developmenttalk/revolutionizing-data-collection-big-data-all-data>.

[29] 中华人民共和国国务院(2015). 促进大数据发展行动纲要[EB/OL]. [2017-07-13]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[30] 祝智庭, 孙妍妍, 彭红超(2017). 解读教育大数据的文化意蕴[J]. *电化教育研究*, (1): 28-36.

[31] 祝智庭, 彭红超(2017). 智慧学习生态:培育智慧人才的系统方法论[J]. *电化教育研究*, (4): 5-14.

(编辑:徐辉富)

Interpreting Educational Data Wisdom

ZHU Zhiting¹, PENG Hongchao² & LEI Yunhe³

(1. School of Open Learning and Education, East China Normal University, Shanghai 200062, China;

2. Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai 200062, China; 3. Putuo Modern Educational Technology Center in Shanghai, Shanghai 200062, China)

Abstract: Big data has received widespread attention in society. There is a huge population of teachers and students in the field of education, and the potential value of educational data is also very large. Due to the increasingly apparent limitations of big data, some concepts, such as small data, all data, fast data, actionable data, relevant data and smart data, are starting to attract the attention of data researchers. From the perspective of interpreting educational data wisdom, this paper clarifies the relationships between the above concepts. We propose the idea that big data and small data constitute all data with both data intensive and intelligence intensive. And we also propose fast data, actionable data, relevant data and smart data as the four new paradigms that highlight the potential value of big data from different perspectives. This paper is expected to provide an idea for the mining and application of educational data.

Key words: data wisdom; big data; small data; all data; fast data; actionable data; relevant data; smart data; dark data