

数据驱动的精准化学习评价机制与方法

□黄涛 赵媛 耿晶 王涵 张浩 杨华利

摘要:综合素质评价是深化教育改革、落实立德树人根本任务的必然要求。信息技术赋能教育教学过程,使得传统教育评价机制向数据驱动的精准化学习评价方向迈进。数据驱动的精准化学习评价是一种不断获取、整合、分析学习过程中的多模态数据,对教学和学习活动的信息流与运行轨迹作出观测、解释和反馈,以发现教育中存在的问题,辅助课堂教学干预的评价方法,包含多场景数据采集、多空间数据融合、精准分析模型构建、分析结果可视化四个关键组成部分。多场景数据采集为学习评价提供数据来源和量化手段支撑,多空间数据融合为学习评价提供统一数据标准,精准分析模型为学习评价提供数据驱动的学习分析与评价方法,分析结果可视化对学习评价提供反馈和应用服务。数据驱动的精准化学习评价框架,应以区块链技术为基础,融合人工智能、云计算、学习分析、情境感知等新兴技术,在多维时空尺度上全面采集学习过程中的多元海量数据,通过全面、系统的统计分析和数据挖掘,从学习者的内在动机到认知发展、情感表现,再到社会综合交互能力,多维度、及时、准确地评估学生的学习状况。当前,探索数据驱动的精准化学习评价,可从构建符合多空间数据特征的综合素质评价框架,完善过程性数据的采集与共享机制,突破情感、认知、动机与社交评价的关键技术,以及研制数据驱动的学习评价平台等方面入手。

关键词:数据驱动;精准化;学习评价;评价机制;评价方法

中图分类号:G434 **文献标识码:**A **文章编号:**1009-5195(2021)01-0003-10 doi:10.3969/j.issn.1009-5195.2021.01.001

基金项目:国家自然科学基金面上项目“面向时空融合的学习者认知诊断理论及关键技术研究”(619770332);中国科学院院士咨询项目“我国科学教育发展战略研究”(2018-Z10-A-025)。

作者简介:黄涛,博士,教授,博士生导师,华中师范大学人工智能教育学部、教育大数据应用技术国家工程实验室(湖北武汉 430079)。赵媛,博士研究生,华中师范大学国家数字化学习工程技术研究中心(湖北武汉 430079);讲师,九江学院(江西九江 332005)。耿晶(通讯作者),博士研究生,华中师范大学教育大数据应用技术国家工程实验室(湖北武汉 430079)。王涵,博士研究生,华中师范大学国家数字化学习工程技术研究中心(湖北武汉 430079)。张浩,博士,副教授,硕士生导师,华中师范大学教育大数据应用技术国家工程实验室(湖北武汉 430079)。杨华利,博士研究生,华中师范大学国家数字化学习工程技术研究中心(湖北武汉 430079)。

2019年7月,中共中央、国务院印发《关于深化教育教学改革全面提高义务教育质量的意见》,提出要“构建以发展素质教育为导向的科学评价体系,强化过程性和发展性评价”(新华网,2019)。2020年10月,中共中央、国务院发布《深化新时代教育评价改革总体方案》,将“改革学生评价,促进德智体美劳全面发展”列为重点内容,提出“促进学生全面发展的评价办法更加多元”的改革目标(新华网,2020)。构建多元、全过程、全要素的学生综合评价,需要改革传统的以纸笔测试为主的学业评价,充分利用信息技术,探究以多种教学场景为背景,基于学生学习行为、认知建构、情感体验、

思维变化、学业表现等多模态数据,实现教学相长的精准化学习评价方式。

一、信息时代的学习评价现状

学习评价经历测量、描述、价值判断时代,逐步向意义建构发展。传统以纸笔测试为主的学业评价,往往侧重测试学生认知方面的知识记忆与理解程度,对学生高阶认知能力如创造力的重视不够,效果也不佳(田爱丽,2020)。对非认知能力如学生的尽责表现、情绪调节、合作能力、思想开放、社会交往等(OECD,2019)的发展情况,传统的纸笔测试更是显得无能为力。而且囿于数据和技术的双

重缺失,传统学习评价与反馈结果呈现出模糊性和延时性等特征,这制约了其价值实现。随着信息技术与教育的深度融合,学习环境、学习内容、学习交互方式发生翻天覆地的变化,学习评价也随之发生变化。依据教学评价场景的不同,本文从大规模学习评价、课堂学习评价和在线学习评价三个方面,对信息时代学习评价的规模化、过程化和精准化等特征进行阐述。

1.大规模学习评价

大规模学习评价(Large-Scale Assessment)是一种通过对特定数量学生群体的整体成就水平和教育相关性因素进行价值判断,来监测学生学习效果和判断教育政策绩效的评价方式(郭元祥等,2014)。与传统教育评价使用经典测量理论模型对学习个体进行比较和评估不同,大规模学习评价更注重对不同学习者群体的能力评价。

国外大规模学习评价中最有影响力的当属国际教育成就评价协会发起的国际数学教育比较研究项目(The Trends in International Mathematics and Science Study, TIMSS)与联合国经济合作与发展组织发起的国际学生评估项目(Program for International Student Assessment, PISA)。TIMSS测评从1995年开始以4年为一个周期,通过测试和问卷,测量不同国家学生数学和科学学习状况,以此了解各国课程目标的实现程度。其最初的目标分析框架由课程内容、表现期望和观点三个方面构成,其中观点涉及学生的态度、兴趣等;自2003年开始,知识和认知维度成为其测评的主要维度(Mullis et al., 2014)。PISA测评自2000年起每3年为一个周期,测评不同国家15岁在校生的科学、数学、阅读等核心素养及相关影响因素,以科学反映学生参与未来社会生活的能力,是对基础教育进行跨国家(地区)、跨文化的评价,为教育教学改进提供有效证据。其测评框架由情景、能力、知识、态度四个指标构成,重点评估学生在知识掌握与运用、认知与思维发展、价值观念与个人能力等方面的发展状况,同时研究影响学生学业成就的因素(OECD, 2014)。

我国近年来也非常重视对学生的大规模学习评价。2007年,我国连续8年开展了义务教育阶段6个学科领域的试点测试,其中5次为全国范围的大规模测试,检测样本包括全国31个省(直辖市/自治区)和新疆生产建设兵团,695个样本县(市/

区)的46万余名学生、11万余名教师和校长。监测结果除了呈现全国学生相关学科领域的总体表现、测查学生综合运用知识能力和解决问题能力外,还关注学生的综合素质和健康成长,对学生的情感态度、课业负担、学校间和区域间的均衡状况进行调查,体现了促进学生全面发展的素质教育理念(中国政府网,2015)。2015年4月,国务院教育督导委员会办公室印发《国家义务教育质量监测方案》,依据我国义务教育课程设置的基本要求,以义务教育阶段四年级和八年级学生为监测对象,利用纸笔测试工具(学科测试卷和调查问卷)和现场测试工具(现场演示和项目参与),监测义务教育阶段学生语文、数学、科学、体育、艺术、德育等学科学业质量、身心健康及变化情况,深入分析影响义务教育质量的主要原因(柴葳等,2015)。

2.课堂学习评价

学习评价最重要的目的不是为了证明,而是为了“改进”(Stufflebeam, 1998)。课堂学习评价的最终目的是改进教学和促进学生学习。传统的课堂学习评价是在教师个人经验的基础上,对学生课堂学习情况做出的主观判断。这种基于个人经验的分析,结果难免受观察人的主观意识的影响。Yamtim等(2014)通过研究发现,课堂学习评价的现状并不乐观,课堂学习评价受到各种因素的限制,其可操作性差。一些研究者提出了操作性改进策略,如日本学者提出在教学中实践“一页纸档案袋评价”(田中耕治等,2012),即鼓励学生将学习经历记录在一页纸上,借此让学生了解个人的学习历程,同时也让教师可以给予适当的指导来帮助学生改进。荷兰的研究者和教师通过紧密合作,共同开发了课堂形成性评价模型(Classroom Formative Assessment, CFA),用于帮助教师判断学生是否达到学习目标并提供及时的反馈(Van den Berg et al., 2016)。还有一些学者旨在开发评价工具来优化课堂评价过程。如瑞士学者Smit等(2017)开发了一款叫Rubies的工具,该工具能够记录课堂师生的互动信息,智能分析与诊断学生的技能状态和能力水平,以此来提升教师对学生学习的感知和诊断能力,同时完善学生自我评价和课堂形成性评价。

随着人工智能等新技术在课堂教学中的广泛应用,学习行为数据呈现出全时段、全方位、多模态的特征,基于过程性数据的学习评价引起了广泛关注,越来越多的研究开始关注学习情绪、学习态

度、学习习惯等多元特征 (Fredricks et al., 2004)。一些研究者提出通过对视频中中学生面部、眼部等特征进行分析,以此来判断学生学习专注程度,进而评估课堂学习效果。如Whitehill等(2014)根据面部表情,进行学生专注度的自动识别;Psaltis等(2017)提出从头部运动、眼睛运动模式中获取线索,推断电子学习环境中的学生学习专注度;Kamath等(2016)使用开源工具库OpenFace和OpenPose提取包括眼睛注视、头部姿势和身体姿势在内的多模态特征,进行学生专注度预测。还有一些学者关注学生学习情绪与成绩的关系,如有研究发现学业情绪与学生的认知和行为密切相关,是影响智力投入、行为持久性以及学习成绩的重要因素(药文静等,2020)。Hsieh等(2019)探讨了在任务转换范式下不同情绪如何调节认知灵活性的问题,发现积极情绪并不能调节短期或长期认知的灵活性,而消极情绪有助于认知的瞬间转换,并导致较小的转换成本。刘君玲等(2019)强调了学业情绪在在线协作问题解决中的重要性,并从环境、活动设计、学习者等方面提出了提高学生协作问题解决能力的策略。Eilam(2019)为检验情绪氛围在课堂中的作用和功能提出了新的分析框架。Ashwin等(2020)通过识别学生的面部表情、手势和身体姿势来分析教室中学生的情感状态。Deniz等(2019)开发了可以通过学生的面部表情监测学生情绪,并进行情绪预警的工具。

3.在线学习评价

信息时代,在线学习已成为人们学习的重要方式。关于在线学习评价,国内外也已展开广泛的研究。虚拟现实技术(Johnson-Glenberg, 2019)、脑波监测(Liu et al., 2016)和眼动(Lambert et al., 2007)等智能感知技术为多模态数据的采集提供了可能。伴随深度学习等人工智能技术快速发展,基于多模态数据的统一表征来开展精准化学习评价已成为在线学习评价的一大趋势。当前在线学习评价的内容,主要包含以下几方面:

学习者认知诊断。随着心理测量学和认知心理学的进一步发展和现代教育技术水平的不断进步,诊断测试的价值和意义不再仅仅是提供分数,而是对学习者的认知结构的测量与评估。基于传统概率图模型的认知诊断技术,可以对学生知识结构进行深入分析。如美国教育巨头McGraw Hill基于人工智能技术开发的学习评估系统(Assessment and Learning in Knowledge Spaces, ALEKS)。该系统基

于知识空间理论,使用人工智能引擎为每位学生绘制个人知识图谱,以此来帮助学习者开展适应性学习(Canfield, 2001)。而基于深度学习方法的知识跟踪技术,还可以获取学生的知识掌握动态,从而分析其学业发展情况。

情感态度分析。学习者的认知加工过程是非常复杂的,存在诸多显性与隐性的相关因素,已有的认知诊断只能诊断显性认知属性(认知行为),即学习者对知识、技能的掌握情况;而对于隐性的相关因素,如心理、情感等,并未加以考虑。然而这些隐性因素也是影响诊断结果的重要因素,忽略了这些因素就容易造成诊断模型和真实认知加工过程之间的不匹配,从而导致结论片面化、经验化。传统教育情境下对学习者的情感态度的分析,通常是通过问卷、对话、观察等方式进行,教师难以有效获得学生的真实情绪状态。计算机视觉技术的发展,使得课堂学习者的情绪监测成为可能。孟菲斯大学研发的AutoTutor智能助手系统,能够感知学习者的情感和注意状态变化(Craig et al., 2013),并根据这些变化对学习者的情感和注意力评价,进而调整对学习者的辅导。

数据驱动的学习预测。利用学生数据来预测学生学习行为和成绩已成为近年来教育领域的一个重要层面(Lodge et al., 2017)。大数据技术可以帮助教师、学生和其他利益相关者洞察学习过程(Jacqueleen, 2015)。如宾夕法尼亚州立大学基于2005-2016年850万条学生数据的分析,构建了数据驱动的高校学生学业表现预测模型,通过综合评估,识别可能面临风险的学生,帮助学校管理部门提前制定干预策略(陈新亚等,2020)。

综上所述,在信息技术的支持下,学习评价的规模化、过程化、精准化等都已取得较大进步,但是也还面临一些问题。如在大规模学习评价方面,由于评价对象数量大、覆盖范围广,使得数据的采集与分析呈现一定的难度。而且,大规模学习评价多以学习者的知识技能、课业负担等显性内容为主,对学习者的综合素养以及学习能力的评价稍显不足。又如在课堂学习评价方面,虽然信息技术嵌入课堂,使得学习评价从单一的结果性评价向关注过程性评价转变,但还存在非认知行为判断标准缺失、学习分析手段和工具还有待进一步多样化等问题。而在在线学习评价方面,虽然深度学习等人工智能方法在学习评价中得到应用,但其评价结果的

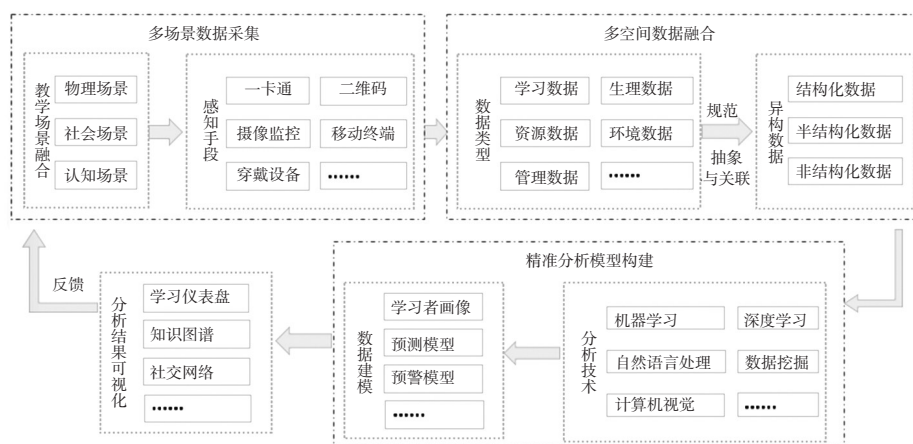


图1 数据驱动的精准确化学习评价机制

解释性问题还需与多元评价方法相结合。

二、数据驱动的精准确化学习评价机制

数据驱动的精准确化学习评价是对多空间、多场景、多时段、多过程的数据进行记录、采集、加工、处理，使得数据在平台间流转累加，实现“1+1>2”的能效的一种评价方法。它由多场景数据采集、多空间数据融合、精准分析模型构建、分析结果可视化等要素组成。多场景数据采集为学习评价提供数据来源和量化手段支撑，多空间数据融合为学习评价提供统一数据标准，精准分析模型为学习评价提供数据驱动的学习分析与评价方法，分析结果可视化表达为学习评价提供反馈和应用服务（如图1所示）。

1. 多场景数据采集

学习空间是开展学习活动的基础环境，也是教育数据产生、应用与迭代流转的重要场所。由于信息技术的介入，学习空间从传统单一的教室实体学习空间向物理空间和网络空间高度融合的融合式学习空间变迁。学习者无论在何种学习空间开展学习活动，都会产生相应的教育场景数据（McAfee et al., 2012）。相较于单维学习空间，融合式学习空间场景较为多元，不仅包括图书馆、教室、阅览室等物理环境，同时也包括角色分配的社会环境以及思维意识介入的认知情境（杨现民等，2020）。由于传统教学场景如学校、图书馆的数字化与智能化水平相对滞后，教学场景中的学习数据采集，主要依赖于人工观察和用户自报告等手段，采集到的数据往往不够全面，且具有一定的主观性（刘智等，2019）。随着可穿戴技术、物联网与人工智能技术的兴起，传统教学场景中的多元化数据得以

采集。研究者可以利用数据感知技术、一卡通、视频监控、智能移动终端、穿戴设备、二维码、无线网络设施等，随时感知与测量学习者的学习数据、生理数据、行为数据、管理数据等并对其进行记录和存储。

2. 多空间数据融合

在评价中，已获取的教育数据并非在同一时刻被分析处理，而是依据评价目标、评价对象，选择不同场景、不同时间节点、不同频次、不同时长、不同维度的数据进行综合分析，因此数据的采集可能会出现数据不统一、噪声干扰、数值缺失等问题。为了将获取的数据转变为计算机能够识别并处理的结构化、半结构化、非结构化数据，为下一步数据建模做准备，需要对已获取的数据进行规范、涉及的关键步骤包括数据清理、数据集成、数据规约和数据变换（刘明吉等，2000）。值得注意的是，一些数据既可以用于建模，也可以用于分析，如学习投入度的评价可采集学习者观看视频时回答问题的频次与中断视频的频次，也可以依据学习者面部表情与坐姿进行分析（Ashwin et al., 2018）；而面部表情同样也可作为学习兴趣的分析方法（孙发勤等，2019）。

3. 精准分析模型构建

分析建模是精准确化学习评价过程中至关重要的一步。其实质是首先依据评价目标的不同，构建相应的评价指标体系；然后对评价指标体系中的评价维度进行数据化表征，采集多源数据；最后基于机器学习、深度学习、自然语言处理、数据挖掘、计算机视觉等技术，对所获取的规范化数据进行建模，最终获得分析结果。分析建模主要包括学习者画像、预测模型与预警模型三类。学习者画像通过

采集到的基础信息、学业数据、学习资源等实时数据,基于文本挖掘、自然语言处理等方法,对学习者的学习特征与个性特征进行描述,帮助教学利益相关者了解学习者学习情况。预测模型是基于学习者持续变化的学业数据、心理数据与生理数据,通过语义关联及相关性分析方法,诊断学习状态并预测未来变化趋势。预警模式建立在预测模型的基础之上,是预测的下一个阶段,对预测的结果进行预警。构建预警模型的主要方法是关联规则、K-means算法等。

4. 分析结果可视化

计算机构建的分析模型较为抽象,不易理解。如果计算机内部自动分析与判别,然后为教师或学习者提供相关学习资源,其分析结果则不需要进行可视化呈现。然而,人在教育中的作用是无法取代的。无论空间如何融合,最终学习数据及学习分析结果都需要方便教师使用。因此精准分析建模后需将分析结果转变为人们易于理解的图形、图像,即将分析结果可视化。可视化分析工具的设计是“数据驱动教学”的核心。当前应用较多的学习仪表盘,被定义为“为了支撑和改进学习和表现,对学习分析结果进行可视化和直观显示的学习分析工具”(Yoo et al., 2015)。它能够直观地呈现学习者学习的相关信息,如学习者对知识点的掌握程度、学习进展情况、登录次数等。知识图谱(Knowledge Graph)是另一个可视化工具,用于绘制、分析和显示学科或学术研究主体之间的相互联系,是揭示科学知识发展进程与结构关系的可视化工具。还有社会网络,能够将学习者群体中沟通与交流的路径可视化,找出意见领袖、边缘人物等。

三、数据驱动的精准化学习评价方法

改进教学和学习的目的,对学习评价的及时性和精准性提出了很高的要求。为了满足新型教育情境下精准评价需求,需要对现有的教育评价理论以及关键技术进行深度剖析,构建数据驱动的精准化学习评价新模式。

针对传统的教学场景,新课标提出了“知识与技能、过程与方法、情感与价值观”的三维教学目标。针对在线教学场景,Fredricks等(2004)提出了行为、认知、情感三维学习评价指标体系;王敏娟构建了以学习者参与为核心的Cybergogy模型,提出学生在线学习效果应涉及认知、情感和社会三

个领域的因素。然而,以上评价框架都忽略了学习动机对学习成就的影响。学习动机是推动学习者学习的直接原因和内部动力,是激发、指引和维持学习行为的心理过程(张雪莲等,2009),因此,学习评价首先需要对学习者学习动机进行评价。其次,无论是在传统的物理学习空间,还是在在线的云学习空间,学习者的学习情况主要是通过认知与情感变化来体现的,故二者是学习评价必须关注的重要指标。最后,传统的行为指标在云学习空间中显得过于笼统,在物理空间与云空间相融合的学习空间中,学习者的社会交互行为等社会因素对学习者的学习成就有非常重要的影响,因此,社会因素分析也是学习评价的重要方面。基于以上分析,本研究建构了内外部因素共同驱动的精准化学习评价框架,即从“动机、认知、情感与社会”四个方面来对学习者的学习情况进行全面的分析与评价,其框架如图2所示。

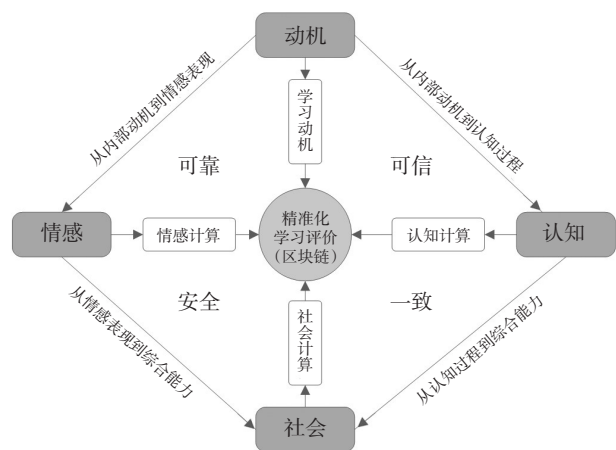


图2 数据驱动的精准化学习评价框架

由图2可知,数据驱动的精准化学习评价框架,以区块链技术为基础,融合人工智能、云计算、学习分析、情境感知等新兴技术,在多维时空尺度上全面采集学习过程中的多元海量数据,通过全面、系统的统计分析和数据挖掘,从学习者的内在动机到认知发展、情感表现,再到社会综合交互能力,多维度、及时、准确地评估学生的学习真实状况。在这个过程中,区块链技术作为核心,可以对学习者数据进行安全存储和隐私保护,使评价结果可信、可靠、安全且一致,为促进学生全面发展提供了支持。

1. 学习动机与精准化学习评价

学习动机是引发与维持学生学习行为,并使之指向一定学业目标的一种动力倾向。传统的学习动

机评价主要通过自我知觉、归因或教师对学生面部表情等的观察来判断。但自我知觉或归因受主观意识影响较大,而教师观察也易受经验和视觉范围等限制。智能感知技术及设备的发展为学习动机的评价提供了便利。如Psaltis等(2017)通过眼动仪采集学生课堂学习的眼动数据,并进行量化和聚类分析,以此来判断学生的注意力焦点,进而获得学习者在学习过程中的专注度情况。Li等(2020)利用神经网络对采集到的学生课堂学习图像数据和学习轨迹进行分析,以判断学生在学习过程中的参与度。Chan等(2020)利用深度学习技术分析学习过程中产生的多模态数据,对学生的情感态度、学业投入、课堂专注等进行量化评价,进而分析学习者的学习动机。

2. 认知计算与精准化学习评价

数据驱动的精准化学习评价模型强调在大数据时代通过“让数据发声”,对学习者的多维度多层次的精准评价,以充分发挥学习评价的诊断、调节、反馈、激励等作用。基于安德森(Lorin W. Anderson)的认知目标分类,学习者的认知由低到高可分为“记忆、理解、应用、分析、评价和创造”等6个层次。数据驱动的精准化学习评价模型不仅要关注学习者的知识掌握情况,还需关注高层次的认知技能;不仅要关注学生的行为表现,还需关注学习者的认知心理建构过程。认知诊断模型是一种基于认知心理学与统计学相结合的认知计算模型(涂冬波等,2012),可以通过对学习者的认知心理过程进行建模,挖掘学生的技能掌握状态。然而,传统的认知诊断模型仅仅是对单次测评的诊断,诊断结果不够稳定。动态认知诊断评价方法对学习者的认知加工过程进行建模,既注重对高阶认知能力的诊断评价,又结合学习者在学习过程中学习与遗忘等认知规律,可以动态跟踪学习者的认知状态变化,通过及时反馈和干预来影响学习者的学习表现。将认知计算与精准化学习评价相结合,可以提高学习评价的动态性、全面性与科学性,为培养学生的高阶思维能力提供帮助。

3. 情感计算与精准化学习评价

情感影响着人的认知过程与行为,情感评价是学习评价的重要组成部分。传统的情感评价方法主要为观察法、访谈法和问卷法,费时费力,且以总结性评价为主,脱离课堂情境,主观性较强。人

工智能技术、视频捕捉技术、视频分析技术的出现为学习过程中的情感评价提供了有力的技术支撑。情感计算是人工智能发展的重要方向之一,主要包括情感识别、情感建模和情感反应三大部分。将情感计算与精准化学习评价结合,可以对学习者在学习过程中产生的多元数据进行捕捉与采集,整合多模态数据,实现对学习者情感状态及变化趋势的实时评价。与音频技术相比,视频分析技术在实时性与持续性捕捉方面具有得天独厚的优势,能够将学习者在学习过程中的面部表情、身体姿态、坐姿等图像进行实时捕捉,结合其他模态数据如脑电波、血压等数据,通过神经网络等技术进行建模与分类,可以追踪学习者学习过程中情绪的变化,精准判断学生的情绪。多模态数据驱动的视频情感评价方法,利用图像识别、语义识别和深度学习技术来量化表征教学视频并构建情感识别模型,自动分析学生在学习过程中的情绪态度,进而推断学生的学习兴趣、学习意志与学习投入度。目前,已开展的相关研究包括通过静态图像对学习者的情绪识别(韩丽等,2017);通过视频分析技术,对传统课堂中学习者的面部表情与姿势进行采集与评价,判断学习者在学习过程中的学习情绪及其变化过程(郭雯雯等,2019)等。还有研究将视频中的表情图片与脑电波的多模态数据集,构建了多模态融合的深度神经网络模型,判断学习者的学习参与度(曹晓明等,2019);或是对视频中的人脸及表情进行识别,评价班级和个人的情感,包括学习兴趣和学习态度、学业价值观、学习动机、学习信念与学习意志(孙发勤等,2019)。

4. 社会计算与精准化学习评价

思维的根本在于对话,对话可以利用社会建构的语境赋予思想以意义(Hudson,2002)。社会互动是教育过程的关键组成部分,学习中的互动是学习者获取知识和发展认知技能的必要条件(Barker,1994)。社会计算是信息技术在社会活动中的应用。从计算技术到社会活动这一角度出发,社会计算就是利用先进的信息技术达到高度有效的交流。基于学习者社交特性的学习者群体评价方法,可以通过分析群体联系的紧密程度和活跃程度来了解群体特性,通过测量群体内部的联结度来分析知识和情感传递的有效性,还可以利用群体权力分析量化学习者在群体中所处的位置。社会计算和精准化学习评价的结合,可以通过对群体层次结构的分

析,帮助我们了解群体内部凝聚力的高低;比较不同学习者群体的互动内容和深度,发现学习者群体之间互动模式的差异(Shu et al., 2018)。

5. 学习者数据的可信存储及隐私保护方法

研究数据的可信存储方法是确保评价真实可靠的基础。采用区块链技术,构建学习者行为链,用以记录学习者过程数据,为精准化学习评价提供可信数据支撑;同时,研究基于区块链的学生评价数据安全与隐私保护技术,以便为多主体评价信息提供可靠的保障,这些都是学习评价参与主体重点关注的问题。面向区块链的访问控制、安全多方计算、同态加密等信息安全手段越来越受研究者青睐。在区块链+教育场景中,参与计算的各方能够保证在不向验证者提供任何有用信息的情况下,使验证者相信某个论断是正确的,因而能够有效避免引入第三方造成对数据的窃取或者窥视,进一步保护教育数据的安全性和隐私性,实现在完全不透露任何隐私信息的同时维持信息的一致性。

四、发展展望与政策建议

目前,数据驱动的学习评价研究在我国仍处于起步阶段。但随着教育与信息技术的深度融合,大数据背景下的学习评价研究呈快速发展趋势。未来,在政策的指引及技术的支持下,探索数据驱动的精准化学习评价大有可为。

1. 构建符合多空间数据特征的综合素质评价框架

智能教育时代,智能感知技术为多空间、多场景和全过程的学习数据采集提供了有效的工具,使得学习过程数据的全面采集成为可能。从数据空间来看,通过智能感知设备,我们不仅可以获得在教室等物理空间学习时的课堂交互数据,同时也可获得在网络学习空间中的在线学习数据。这些数据覆盖课程、作业、生活、社交等多个学习场景,各个场景有各自的评价标准。同时这些学习数据具有多模态特性,包括视频、语音、文本和图片等,亟需归一化标准和通用的评价框架。因此,精准化学习评价亟需建立符合多空间全数据的综合评价框架和指标体系,用于汇聚来自不同场景、不同模态的学习评价数据,为后续大数据精准分析做好准备。

2. 完善过程性数据的采集与共享机制

学生综合素质的评价需要全面采集学生德育、智育、体育、美育、劳动教育等多维过程性数据,

数据来源场景多样、涉及的信息维度多、模态不一。因此,完善国家、区域、学校和班级等多级资源与数据的采集与共享机制,是实施数据驱动的精准化学习评价的前提。人工智能技术与感知设备的快速发展为完善过程性数据的采集奠定了技术基础。研究者可以通过多种传感器设备实时记录学生行为数据,同时,结合观察、访谈、资料分析等不同方法,多维度、多层次地采集学习者的学习结果数据和学习过程数据。如可以通过手环等传感器设备来实时感知学生的运动、位置、睡眠状态等生理数据,通过摄像机等设备自动采集学生课堂学习行为数据,通过个人档案、问卷调查等方式记录学生综合发展数据,通过阶段性测评与诊断性测评采集学生学业数据等等。同时,研究者还要注意完善真实的情景化任务、成果、行为等多源、异构、多模态、不连贯语义的海量数据的共享。不同教育机构的数据与资源共享机制是保证学习评价大数据的完整性与全面性的基础。

3. 突破学习评价中情感、认知、动机和社交评价方面的关键技术

当前学生的综合素质评价大多采用填报与统计方式,其评价的效度和真实性受到质疑。而传统的评价模型侧重学习者的知识掌握情况,忽略了更高层次的知识迁移能力、学习方法、情感态度和社会交往能力的发展情况。这既是重结果、轻过程的评价理念所致,同时也受到评价技术的制约。要实现精准化的学习评价,必须在情感、认知、动机和社交评价的关键技术方面有所突破,通过采集文字、语音、图像、视频等包含认知、情感、注意力等的多模态数据,利用自然语言处理、图像识别、情感、计算等深度学习技术和统计建模方法,分析学习者的学习过程、学习内容、知识迁移、情感获得、交流互动等,以此对学生综合素质进行全面、精准、多维度的评价与反馈。

4. 研制数据驱动的学习评价平台,加快学习评价在教育教学中的应用

数据驱动的学习评价平台需要云计算、数据仓库、数据挖掘等技术和应用的配合,实现低成本的大规模分布式存储、及时响应的用户体验,以及更加绿色节能的新一代数据中心。即在有效面对大数据处理需求的同时,实现最大化资源利用率、最小化系统能耗的目标。为此,数据驱动的学习评价平台需要为教育大数据的汇集融合、海量存储、高效

计算、深入分析挖掘以及教育应用的开发与运行提供基础支撑。基于这个原理搭建教学、管理与学习的全生态应用平台,构建不同业务系统共享数据中心的聚合型平台。在规范化的数据基础之上构建特色型分析模型满足分析需要,为管理者、教师、学生、家长以及合作学校提供智能服务。学校需将校内与校外、课内与课外、线上与线下、网络空间与物理空间产生的数据汇集到学习评价平台中,通过数据不断的流转迭代,帮助管理部门、教师、学生、家长等利益相关者了解学生的学习情况,开展精准的教育干预与辅助。教师应不断提升信息化教学能力与数据素养,能够对学习者不同学习时段产生的数据进行分析,基于相关数据开展精准化教学及互动;家长应努力提升育儿专业知识,学会客观看待学习者产生的相关数据并提供个性化的辅助;学校管理者应提升信息化领导力,能够依据年级数据或全校性数据进行科学归因与精准决策,促进学校教学资源的均衡配置与教学效能的提升。

五、结语

大数据时代,学习评价作为教育生态中的一环,对实现人的德智体美劳全面发展起着非常重要的作用。数据驱动的精准确学习评价,通过不断获取、整合、分析学习过程中的文字、声音、图像、视频与人体感知等多模态数据,对教学和学习活动的信息流与运行轨迹作出观测、解释和反馈,帮助教育相关者发现教育中存在的问题,辅助课堂学习中个性化干预的实现(朱晓玲,2019)。本文基于对信息时代学习评价现状的分析,构建了包含多场景数据采集、多空间数据融合、精准分析模型构建、分析结果可视化的精准确学习评价机制,并提出以区块链技术为核心,从认知、情感、动机和社会四个方面进行学习综合素质评价的方法,为促进教育教学的改进与学业质量的提升提供了借鉴。

总之,相对于传统结果性教育评价,数据驱动的精准确学习评价从多维度关注学习者的学习过程,能够促进教育管理者和教师对学习者的理解,为实施精准化的教学干预提供辅助。不过,要建立一个全面、可靠、真正对学习者有效的学习评价机制需要教育教学论专家、学校管理者和技术专家等多方协作。

参考文献:

- [1][日]田中耕治(2012).日本形成性评价发展的回顾与展望[J].项纯.全球教育展望,41(3):3-6,18.
- [2]曹晓明,张永和,潘萌等(2019).人工智能视域下的学习参与度识别方法研究——基于一项多模态数据融合的深度学习实验分析[J].远程教育杂志,37(1):32-44.
- [3]柴葳,向倩芸(2015).我国建立义务教育质量监测制度,今年起开展监测工作——中国特色“PISA”揭开面纱[EB/OL].[2020-11-07].http://www.moe.gov.cn/jyb_xwfb/xw_fbh/moe_2069/xwfbh_2015n/xwfb_150415/150415_mtbd/201504/t20150416_187197.html.
- [4]陈新亚,李艳(2020).《2020地平线报告:教与学版》的解读及思考——疫情之下高等教育面临的挑战与变革[J].远程教育杂志,38(2):3-16.
- [5]郭雯雯,杨凤梅(2019).课堂环境双模态情感评价系统设计与效果分析[J].教学研究,42(3):27-33.
- [6]郭元祥,刘晓庆(2014).大规模学业评价的发展历程、新趋势及启示[J].教育研究与实验,(1):27-32.
- [7]韩丽,李洋,周子佳等(2017).课堂环境中基于面部表情的教学效果分析[J].现代远程教育研究,(4):97-103,112.
- [8]刘君玲,刘斌,张文兰(2019).学业情绪对在线协作问题解决的影响研究[J].中国电化教育,(7):82-90.
- [9]刘明吉,王秀峰,黄亚楼(2000).数据挖掘中的数据预处理[J].计算机科学,27(4):54-57.
- [10]刘智,方常丽,刘三妍等(2019).物理学习空间中学习者情绪感知研究综述[J].远程教育杂志,37(2):33-44.
- [11]孙发勤,邓雯心(2019).基于人工智能的课堂学习情感评价研究[J].中国教育信息化,(23):58-62.
- [12]田爱丽(2020).综合素质评价:智能化时代学习评价的变革与实施[J].中国电化教育,(1):109-113,121.
- [13]涂冬波,蔡艳,丁树良(2012).认知诊断理论、方法与应用[M].北京:北京师范大学出版社.
- [14]新华网(2019).中共中央 国务院关于深化教育教学改革全面提高义务教育质量的意见[EB/OL].[2020-11-07].http://www.xinhuanet.com/politics/2019-07/08/c_1124725254.htm.
- [15]新华网(2020).中共中央 国务院印发《深化新时代教育评价改革总体方案》[EB/OL].[2020-11-07].http://www.xinhuanet.com/2020-10/13/c_1126601551.htm.
- [16]杨现民,李怡斐,王东丽等(2020).智能时代学习空间的融合样态与融合路径[J].中国远程教育,(1):46-53,72,77.
- [17]药文静,姜强,李月等(2020).众包知识建构下学业情绪影响因素及唤醒机制的科学学研究——面向深度学习的课堂教学结构化变革研究之五[J].现代远程教育,(5):33-42.
- [18]张雪莲,高玲(2009).学习动机及其相关研究[J].教育理论与实践,29(18):32-34.
- [19]中国政府网(2015).教育部督导办负责人就《国家义务教育质量监测方案》答记者问[EB/OL].[2020-08-11].http://www.gov.cn/xinwen/2015-04/15/content_2847308.htm.

- [20]朱晓玲(2019).小学数学课堂学习评价的现状 & 改进策略研究[D].南京:南京师范大学.
- [21]Ashwin, T. S., & Guddeti, R. M. R. (2018). Unobtrusive Students' Engagement Analysis in Computer Science Laboratory Using Deep Learning Techniques[C]// Proceedings of 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT). IEEE: 436-440.
- [22]Ashwin, T. S., & Guddeti, R. M. R. (2020). Automatic Detection of Students' Affective States in Classroom Environment Using Hybrid Convolutional Neural Networks[J]. Education and Information Technologies, 25(2): 1387-1415.
- [23]Barker, P. (1994). Designing Interactive Learning[M]. Dordrecht: Springer: 1-30.
- [24]Canfield, W. (2001). ALEKS: A Web-Based Intelligent Tutoring System[J]. Mathematics and Computer Education, 35(2): 152-158.
- [25]Chan, M. C. E., Ochoa, X., & Clarke, D. (2020). Multimodal Learning Analytics in a Laboratory Classroom[M]. Cham: Springer: 131-156.
- [26]Craig, S. D., Hu, X., & Graesser, A. C. et al. (2013). The Impact of a Technology-Based Mathematics After-School Program Using ALEKS on Student's Knowledge and Behaviors [J]. Computers & Education, 68: 495-504.
- [27]Deniz, S., Lee, D., & Kurian, G. et al. (2019). Computer Vision for Attendance and Emotion Analysis in School Settings [C]// 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). IEEE: 134-139.
- [28]Eilam, E. (2019). Synchronization: A Framework for Examining Emotional Climate in Classes[J]. Palgrave Communications, 5(1):1-11.
- [29]Fredricks, J. A., & Paris, B. A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence[J]. Review of Educational Research, 74(1): 59-109.
- [30]Hsieh, S., & Lin, S. J. (2019). The Dissociable Effects of Induced Positive and Negative Moods on Cognitive Flexibility [J]. Scientific reports, 9(1):1-10
- [31]Hudson, B. (2002). Critical Dialogue Online: Personas, Covenants, and Candlepower[M]. London: Sage: 53-90.
- [32]Jacqueleen, A. R. (2015). The Skinny on Big Data in Education: Learning analytics Simplified[J]. TechTrends, 59(2): 75-79.
- [33]Johnson-Glenberg, M. C. (2019). The Necessary Nine: Design Principles for Embodied VR and Active Stem Education [M]. Singapore: Springer: 83-112.
- [34]Kamath, A., Biswas, A., & Balasubramanian, V. (2016). A Crowdsourced Approach to Student Engagement Recognition in E-Learning Environments[C]// Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE: 1-9.
- [35]Lambert, A. D., Terenzini, P. T., & Lattuca, L. R. (2007). More Than Meets the Eye: Curricular and Programmatic Effects on Student Learning[J]. Research in Higher Education, 48(2): 141-168.
- [36]Li, Z., & Zhan, Z. (2020). Integrated Infrared Imaging Techniques and Multi-Model Information via Convolution Neural Network for Learning Engagement Evaluation[J]. Infrared Physics & Technology, 109: 103430.
- [37]Liu, C. J., & Huang, C. F. (2016). Innovative Science Educational Neuroscience: Strategies for Engaging Brain Waves in Science Education Research[M]. Singapore: Springer: 233-247.
- [38]Lodge, J. M., & Corrin, L. (2017). What Data and Analytics Can and Do Say About Effective Learning [J/OL]. NPJ Science of Learning. <https://www.nature.com/articles/s41539-017-0006-5>.
- [39]McAfee, A., Brynjolfsson, E., & Davenport, T. H. et al. (2012). Big Data: The Management Revolution[J]. Harvard Business Review, 90(10): 60-68.
- [40]Mullis, I. V., & Martin, M. O. (2014). TIMSS Advanced 2015 Assessment Frameworks[DB/OL]. [2020-12-05]. <https://files.eric.ed.gov/fulltext/ED559364.pdf>.
- [41]OECD (2014). PISA 2012 Results: Creative Problem Solving: Students' Skills in Tackling Real-Life Problems (Volume V) [DB/OL]. [2020-08-01]. <http://www.oecd.org/pisa/key-findings/PISA-2012-results-volume-V.pdf>.
- [42]OECD (2019). Social and Emotional Skills Well-Being, Connectedness and Success[EB/OL]. [2020-08-01]. [https://www.oecd.org/edu/school/UPDATED%20Social%20and%20Emotional%20Skills%20-%20Well-being,%20connectedness%20and%20success.pdf%20\(website\).pdf](https://www.oecd.org/edu/school/UPDATED%20Social%20and%20Emotional%20Skills%20-%20Well-being,%20connectedness%20and%20success.pdf%20(website).pdf).
- [43]Psaltis, A., Apostolakis, K. C., & Dimitropoulos, K. et al. (2017). Multimodal Student Engagement Recognition in Prosocial Games[J]. IEEE Transactions on Games, 10(3): 292-303.
- [44]Shu, H., & Gu, X. (2018). Determining the Differences Between Online and Face-to-Face Student - Group Interactions in a Blended Learning Course[J]. The Internet and Higher Education, 39:13-21.
- [45]Smit, R., Bachmann, P., & Blum, V. et al. (2017). Effects of a Rubric for Mathematical Reasoning on Teaching and Learning in Primary School[J]. Instructional Science, 45(5): 603-622.
- [46]Stufflebeam, D. L. (1998). Conflicts Between Standards-Based and Postmodernist Evaluations: Toward Rapprochement[J]. Journal of Personnel Evaluation in Education, 12(3): 287-296.
- [47]Van den Berg, M., Harskamp, E. G., & Suhre, C. J. M.

(2016). Developing Classroom Formative Assessment in Dutch Primary Mathematics Education[J]. Educational Studies, 42(4): 305–322.

[48]Whitehill, J., Serpell, Z., & Lin, Y. C. et al. (2014). The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions[J]. IEEE Transactions on Affective Computing, 5(1): 86–98.

[49]Yamtim, V., & Wongwanich, S. (2014). A Study of

Classroom Assessment Literacy of Primary School Teachers[J]. Procedia–Social and Behavioral Sciences, 116: 2998–3004.

[50]Yoo, Y., Lee, H., & Jo, I. H. et al. (2015). Educational Dashboards for Smart Learning: Review of Case Studies[M]// Emerging Issues in Smart Learning. Springer, Berlin, Heidelberg: 145–155.

收稿日期 2021-01-05 责任编辑 汪燕

Evaluation Mechanism and Method for Data-Driven Precision Learning

HUANG Tao, ZHAO Yuan, GENG JING, WANG Han, ZHANG Hao, YANG Huali

Abstract: Comprehensive quality evaluation is an inevitable requirement for implementing the fundamental task of deepening education reform and cultivating talents. Information technology promotes the process of education and teaching, which makes the traditional education evaluation mechanism move towards the direction of data-driven learning evaluation. Data-driven precision learning evaluation is an evaluation method that observes, explains and feedbacks on the information flow and running track of teaching and learning activities by continuously obtaining, integrating and analyzing multi-modal data in the learning process, so as to find problems existed in education and assist the intervention of classroom teaching. It contains multi-scene data collection, multi-spatial data fusion, accurate analysis model construction, and analysis result visualization. Multi-scene data collection provides data sources and quantitative means for learning evaluation; multi-spatial data fusion provides unified data standards for learning evaluation; accurate analysis model provides data-driven learning analysis and evaluation methods for learning evaluation and analysis results visualization provides feedback and application services for learning evaluation. The data-driven precision learning evaluation framework should integrate technologies such as artificial intelligence, cloud computing, learning analysis and context awareness based on the block chain technology. It can comprehensively collect multiple massive data in the learning process on multi-dimensional spatial-temporal scale. Through comprehensive and systematic statistical analysis and data mining, it can evaluate students' learning status from their intrinsic motivation, cognitive development, emotional performance, to comprehensive social interaction ability in a multi-dimensional, timely and accurate way. At present, to explore data-driven accurate learning evaluation, we can carry out research on building a comprehensive quality evaluation framework in line with the characteristics of multi-spatial data, improving the collection and sharing mechanism of process data, breaking through the key technologies of emotion, cognition, motivation and social evaluation, and developing a data-driven learning evaluation platform.

Keywords: Data Driven; Precision; Learning Evaluation; Evaluation Mechanism; Evaluation Method