

区域性教育大数据总体架构与应用模型*

余胜泉, 李晓庆

(北京师范大学 未来教育高精尖创新中心, 北京 100875)

摘要:教育大数据作为未来区域教育发展的契机和质量改进与提升的破解方法,在智慧教育领域有着不可估量的发展潜能。该文大数据与区域教育发展的关系出发,剖析了区域教育大数据的技术架构,包括教育过程多模态数据收集、学习者个性化认识模型构建、学科知识图谱构建、数据挖掘分析、资源语义标记与汇聚、个性化智能推荐引擎、区域教育决策分析等,提出构建区域大数据无缝流转的开放生态系统。同时,该研究以智能教育大数据公共服务平台“智慧学伴”为例,提出区域教育大数据应用模型,并展望了区域教育大数据未来发展的三个智能服务阶段:个性化定向学习支持的“学习助手”、封闭性问题解答和情感支持的“学习伙伴”、开放性问题解答和智慧成长支持的“学习导师”。该研究从理论构建、实践应用、未来展望三个层面对区域教育大数据作了较为全面的论述,以期为区域教育大数据的应用提供理论参考模型和实践指导。

关键词:区域教育大数据;区域教育质量提升;教育大数据应用;教育大数据生态系统

中图分类号: G434 **文献标识码:** A

一、大数据与区域教育发展

(一)大数据是区域智慧教育发展的基础

智慧教育是当前教育信息化的发展方向,也是区域教育信息化的美好期望,但要实现教育中的智能支持,大数据是必不可少的。近几年以深度学习为代表的人工智能实现了突破,关键在于一方面是大规模并行计算能力的发展,另一方面是大数据的汇聚,通过深度人工智能神经网络学习大数据中蕴含的规律,从而获得了智能,如语音自动识别、语言精准翻译、计算机视觉等。在教育领域,如果要全面了解学生,必须全面采集学生学习过程的各种大数据,通过数据拟合,形成模型。区域性教育大数据汇聚应用是智慧教育必经的前置阶段,依托大数据,逐步形成智能系统和智能装备,为智慧教育发展奠定基础。

(二)大数据是科学配置教育资源的基础

教育大数据是在教育活动运行过程中的数据,它反映教育系统运行过程中实时、真实的运行状况,蕴含着教育系统运行的内在规律。通过数据挖掘,可以将这种规律及其演化趋势可视化展现出来,使教育决策部门可以预测到区域内教育发展的需求趋势,从而进行科学的教育决策与教育资源

配置,由事后补救转向事前预警,使得教育资源配置过程更迅速,甚至达到实时与即时性。如基于区域性的教育电子地图,叠加区域综合大数据,及时可视化动态显示变化趋势,帮助教育决策者获得超越个体与局部的宏观或中观的洞察力;通过数据聚类、关联分析和演化趋势等技术分析,可以发现群体共性问题,群体中的异常模式,相比之前更容易发现问题所在、可能弱点和盲区,并发现背后根源性的问题。

(三)大数据是改进教育公共服务模式的重要契机

教育公共服务向来优先满足群体的共性需求,难以顾及个体的个性需求。究其原因,传统手段很难实现精准了解每个个体的个性化需求,要为每个个体提供个性化服务,成本太高,基本无法实现,而基于互联网和教育大数据来提供教育公共服务,则可以高效解决这个问题。通过对学习数据的深度挖掘与分析,可以让教师了解每个学生的学习特征与需求,为学生推送适宜的学习资源与服务,不需要大规模人力投入,即可实现对个体的实时监测与及时反馈^[1]。互联网可以最大程度汇聚社会教育服务资源,通过大数据来提供精准的导向,提升个体

* 本文系教育部哲学社会科学研究重大课题“‘互联网+’教育体系研究”(项目编号:16JZD043)阶段性研究成果。

获得感。通过建立共性需求与个性需求相兼容的教育公共服务体系,可以构建起具有“更多获得感”的教育公共服务供给体系^[2]。

(四)大数据是推进区域教育质量监测与管理的最佳手段

传统教育质量监测数据是阶段性的、相对静态化、抽样性的数据,一般在于凸显学生群体的发展情况,如学生整体的学习成绩水平,学习负担、心理素质与体质健康等。数据通过周期性、阶段性的监测评估来采集,采集的对象一般采用抽样的方式决定参与的群体。而教育大数据是全样本、全过程的数据,它可以更好地关注每一个学生的微观个性表现,是学生学习过程中产生的数据,如学习的时间、学习内容、学习活动、人际交互、学习投入与情感投入等各类数据,这些数据在日常教育业务中产生,全样本覆盖,实时收集、实时分析、实时反馈、实时改进,数据不仅仅是服务于区域教育行政部门,更可以服务于每一个老师、每一个学生、每一个家长,让服务主体有获得感,是推进区域教育质量改进与提升最有效的手段。

二、区域教育大数据的技术架构

区域教育大数据平台是推进区域性应用的核心关键,大数据平台不仅仅是计算资源虚拟化、云化,更重要的是利用人工智能领域的深度学习、自然语言处理、知识图谱、知识推理、语义搜索、智能推荐等关键技术建立面向教育业务的大数据收集、处理、分析及应用模型。

在良好的大数据生态体系的支持下,需要针对典型教育场景及跨场景的教学环境,进行学习全过程数据的采集、对学习资源进行语义描述与关联。以采集的全体学习者数据为对象,基于认知、情感与动机等关键特征,利用知识跟踪、深度学习等技术构建学习者常模与个性化模型。通过个性化学习服务工具与精准推荐服务引擎,基于海量学习资源

的汇聚,形成感知个性化需求的智能化教学环境,支持开展围绕课内课外、线上线下相结合的个性化教与学的关键环节与途径。在服务的过程中,数据自然的汇聚与生成,形成区域教育大数据,可为区域教育质量管理、教育资源配置、区域性精准教研、区域性教育公共服务提供科学的决策基础。典型区域大数据的技术架构如图1所示。

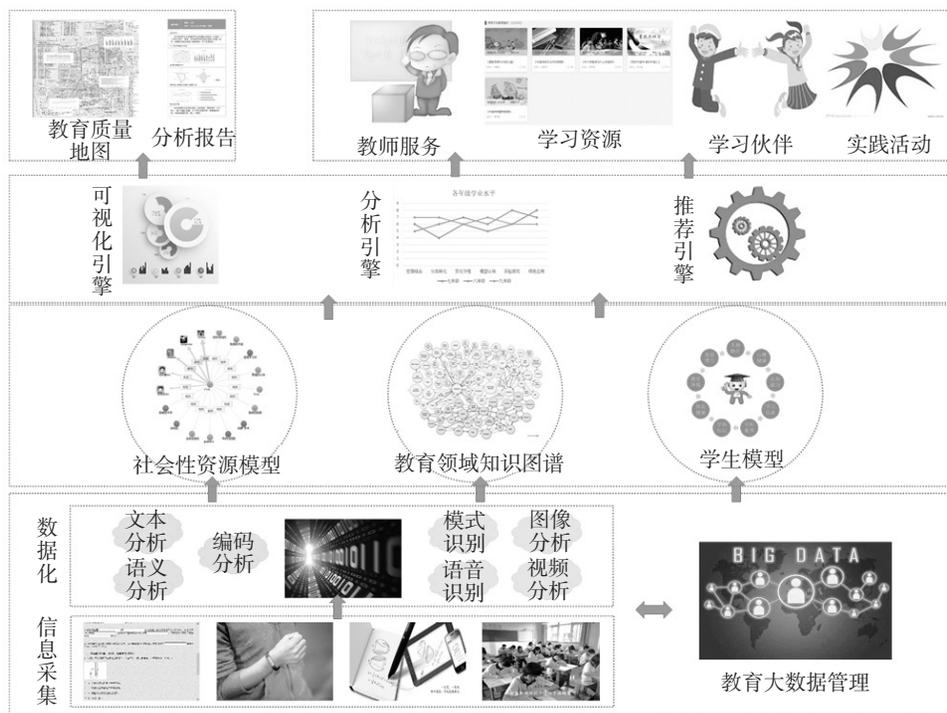


图1 区域大数据平台的逻辑结构

区域教育大数据平台应具备情境感知、自然交互数据全记录、学习资源语义化汇聚、内容自动转码与格式自适应等特性,利用平台中建立的学科知识本体库、认知模型库以及推理引擎等核心基础模块,对采集的学习过程交互数据进行自动化分析,实现学习资源与学习服务个性化、适应性推送。

(一)教育全过程多模态数据的采集

学习过程数据的采集,可以从学习情境信息、知识建构行为信息、学习行为和学习结果信息四个方面来进行。其中学习情境信息主要包括学习者初始能力水平、学习终端等影响学习者获取和运用知识的一切要素。知识建构信息主要包括学习内容和学习资源的编辑、审核、批注、分享和传播等对学习内容和学习资源进行再创造的数据。学习行为信息以学习活动为核心,主要包括浏览学习内容、参与学习活动、完成学习任务等过程性的行为数据。学习结果信息主要包括学习时长、测试成绩等成果性数据^[3]。学习情境数据的采集可以借

助智能可穿戴设备与智能传感器等工具实现,学习行为数据的采集可以通过插码工具和xAPI^[4]等实现,知识建构和学习结果等认知投入、目标状态数据可以借助大规模在线课程中所设计的学习活动进行采集。

教学场景数据采集包括课堂视频实录数据、语音实录数据、各类管理系统记录的数据、师生互动的数据、在校园活动的轨迹数据等。

所采集的跨场景、多种类的异构过程化教学信息,还需要通过文本分析、编码分析、语义分析、语音识别、图像分析等技术,转换成可以挖掘和分析的数据,并基于数据集成交换标准进行融合与管理,以实现教育全过程数据的实时收集、有机集成和全周期管理。

(二)学习者个性化认知模型构建

学习者学习个性特征的刻画是开展个性化支持服务的关键。在学习过程中学习者的个性特征,不能只是网络访问行为等外部特征,更要关注其发展特征与认知特征,包括认知能力、学科能力、问题解决能力、心理状态、体质健康、学习风格、学习情绪与知识的学习交互关系等,需要基于教育与心理的模型来对学习者的个性特征进行语义建模。

在学科能力表现方面,结合学科特点,构建语文、数学、物理、外语(英语)、生物、化学、道德与法治、地理、思想政治、历史、科学、体育、自然、综合实践、音乐、美术等学科的评测与诊断模型,以学科素养—关键能力—核心知识为关键表征与统一的编码体系,形成对各个学科的核心素养与学科能力表现的表征、评测与诊断整体框架。各学科根据本学科的核心素养和内容主题,采用严谨的程序规范,开发关于学生学科素养和学科能力认知的诊断工具,将学科核心素养、知识内容与能力表现密切关联,建立基于多级评分的、考虑学生不同作答策略的、能够表征连续心理特质与离散知识状态的、可以分析群体知识水平的认知诊断模型。

在心理健康特征表征方面,包括对个体的情绪、自我认知、社会适应等多个方面建模。心理健康可分为正负两个方面,既要关注积极的心理健康状况,也要关注消极的心理行为问题。在强调学生情绪、行为之外,还强调教育活动对学生良好人际关系、人格品质、合理社会认知的养成。要基于一系列高质量的综合心理素质测评工具,建立一套中小学生学习综合素质测评常模。

在体质健康表征方面,基于体测数据及日常运动行为数据建立学生身体健康素养综合评定指数体系。通过跟踪诊断学生体质健康、身体运动功能、

运动技能学习、生活方式等,判断体质健康、身体功能及运动技能学习的发展优势。基于数据分析生成科学的体质测试分析报告,帮助家长了解孩子的体质状况和健康状况,关注、提升孩子在科学运动方面的兴趣,引导孩子健康成长,全面发展。

在学习行为特征方面,以大规模学习者样本数据为基础,对不同年龄、性别、区域的学习者群体所表现的学习特征及其典型变异进行分析。基于学习情境信息、学习结果信息、学习行为和知识建构行为四个方面数据对学习者的个性特征进行语义建模。从学习兴趣、学习风格和知识水平等多维度构建学习者常模,形式化、明确化表征学习者个性和共性特征。同时通过对学习者的持续追踪,分析学习行为与知识单元的多尺度关联特征,发现学习者在线学习的共性和个性变化规律,实现对学习者发展常模的动态生成和持续更新。

(三)基础教育学科知识图谱构建

知识图谱通常指利用多关系图结构描述真实世界或特定领域的各类实体以及这些实体间的关联的知识库。教育知识图谱是针对教育领域构建的知识图谱,旨在表示教学过程中涉及的不同元素及这些元素之间存在的具有教育意义的各类认知关系。要构建基础教育领域的知识图谱需要深挖教育全过程中教学、学习、管理、评价四大有机环节中产生的数据,以知识内容、教师、学生、管理者为核心对象,基于角色、任务、行为、效果展开语义抽取、融合和推理,构建动态教育知识图谱。

学科知识图谱可以在教育语义计算、教育资料的语义关联、学习者模型建立、网络教学平台开发、智能答疑系统开发、学习资源的个性化服务等诸多方面发挥重要的基础性作用,是这些应用得以实现的前提。以语文、英语、数学、物理、化学、生物、历史、地理等学科课程标准与教材为基础,建立教育领域学科知识图谱,综合应用人工智能技术辅助学科概念的提取、概念定义性质的提取以及概念关系的发现等,实现领域图谱构建过程的半自动化。

学科知识图谱构建的核心思路是基于知识之间的语义关系所形成知识之间的逻辑关联网,并叠加教学目标、常见教学问题等教学知识。在知识图谱的基础上,叠加学习者对知识掌握的状态信息,用于可视化表征学习者动态知识发展的情况,从而形成学习者的认知地图。认知地图既可以用于知识导航,又可以作为评估的有效支架,还能用于学习内容与学习服务的精准推荐,同时满足用户不同层面的需求^[5]。

(四)教育数据挖掘分析

在数据仓库和知识图谱的基础上,通过教育大数据分析系统,对教育数据展开实时挖掘与分析,包括教学全过程的量化评估与效果预测、学生学习全过程的量化评估与效果预测、学生意外精神状态的预警等,实现知识内容和信息在不同对象之间的有效传递,实现教育大数据与教育服务的无缝融合与价值提升。基本技术思路为:(1)分析教育各环节产生的数据特征,制定面向教育大数据的规范标准,对数据进行清洗、整合;(2)借鉴ESB(Enterprise Service Bus,企业服务总线)的数据集成思路,利用NoSQL、Hadoop、Spark等面向分布式、流式计算的数据存储与分析系统,构建集成的异构教育大数据平台;(3)利用数据分析与知识计算的相关理论、技术,针对教育大数据的特点设计聚类、关联、分类等模型与算法,实现语义发现与推理,提升教育大数据的价值;(4)通过可视化技术,实现基于大数据的智能教育服务的展示、交互等功能。

在业务场景的分析上,关键是要围绕学科教学的发展进程来建模分析:(1)通过对教师、学生的在线行为进行挖掘与分析,描绘学科教学模式、学生交互模式等方面的发展进程,发现热点问题,预测发展方向;(2)对教师之间的合作关系、教研生态等进行分析,发现网络教学团队建设的特点,通过教师线上交互数据的挖掘,发现和发掘优秀教育成果,将专家主导的教育发展模式转变为专家—教师互动共建的模式;(3)对学生最邻近发展区进行表征与计算,对学生动态发展的能力进行评估,分析学生学习的优势和不足,融入教育大数据的层级性、时序性和情境性特征(学习影响因素),构建新一代的认知诊断模型;(4)建立学生心理、体质健康等关键领域常模,结合标准划定技术以及潜类别模型等,有效描述学生的心理与体质健康的离群状态,及早干预,促进学生身心健康发展;(5)汇聚群体数据进行深度挖掘,通过对数据聚类、分类、离群点、关联、相关性、判别、比较、偏差等参数的分析,了解群体的整体水平与发展趋势,形成面向不同群体的分析报告^[6]。

(五)资源的语义标记与聚合

资源要以教师和学生的实际获得最大化为导向,设计、开发和汇聚符合在线学习特征的数字资源与创意服务,建立平台资源与服务的统一标准,提供与外部数字资源与学习服务融合的接口,汇聚海量资源形成“人人教、人人学”的教育服务超市。

教育资源语义化是实现智能推荐的基础。要探

索学习资源的自动化语义标注机制,实现海量学习资源的自动化、精准化语义标注。具体技术路线是对采集的多形态资源数据进行基于结构和内容的语义提取和语义映射处理,实现机器自动化的语义聚合。语义提取过程中针对结构化的开放教育资源如DBpedia、Freebase等采用语义查询语言(如SPAQL)直接获取;对于半结构化的教育资源采用语义包装器抽取资源结构信息,利用自然语言处理技术、概率图模型、机器学习等技术提取资源内容语义信息;对于无结构化的教育资源直接利用自然语言处理技术、神经网络模型等技术提取资源语义信息。在语义映射处理过程中,构建多学科教育资源的桥本体,对通过语义提取获取的教育资源语义信息进行本体对齐处理,保证多源异构教育资源的语义一致。

网络资源的聚合是指对分散在互联网上的资源根据一定的规律进行筛选、排序和重新组合,从而为用户提供更有针对性、更易查找和使用的资源集合。资源聚合的工作原理是对资源进行标识,通过对用户需求分析,整合相关资源实现资源有组织的重构,从而向用户提供精确有效的资源集合。聚合的核心是解决学习资源内在关系的建立以及外在可视化的问题,以便更好地重用与共享。学习资源深度聚合,首先从计算资源的语义关系出发。采用基于规则的推理、语义相似度的计算等技术,动态发现资源之间的各种语义关系,将异构学习资源的语义空间整合到统一的语义空间,深入揭示学习资源的内在联系,并将资源和服务进行深度聚合。其次,促进学习资源深度聚合,也能够利用社会计算的方法,通过分析用户与学习资源的交互行为,从中发现社会关系、社会行为的规律,利用社会化的行为来改善、提升资源聚合的质量和结构化程度。在线学习中,学习者不仅是从内容本身中获得知识进行学习,内容背后的人在学习者学习过程中也扮演着重要角色^[7]。在线学习中,由于知识之间的内在关系与人的社会性交互,形成了人与知识共生的社会性知识网络,社会性知识网络对于学习资源的适应性推荐至关重要。

(六)个性化智能推荐引擎

个性化学习智能推荐引擎是区域大数据平台的核心组成之一,该引擎可以为教师、资源提供者以及学习者提供精准且适当的资源组织形式和获取途径,提供完整的个性化学习支持,包括学习内容和活动推荐、学习工具与策略推荐、学习专家与同伴推荐等,实现个性化学习服务推送。

学习资源与学习服务的智能推荐引擎要借鉴

双向适应的原理,从学习者模型和学习资源组织两个方面的相互联动角度,实现基于数据驱动的、感知学习过程情境的、持续性的自适应。基于数据驱动的自适应关键在于研究如何从用户兴趣、已有知识、动态能力和学习风格等角度出发进行课程资源推荐,并基于内容相似度和链接结构相似度进行目标资源的筛选和聚类。感知学习过程情境的自适应关键在于研究如何获取和表征学习者当前所处的情境信息,并与资源情境本体进行适应匹配,实现对学习内容、学习策略、学习活动、知识专家、学习路径和学习服务的推荐。持续性的自适应关键在于研究如何利用学习过程中的核心数据完成对学习者在在线课程知识节点学习状态的计算,实现对知识掌握水平的客观、动态、连续表征。同时,由于学习是一种持续发展的过程性活动,因此在持续的自适应中需要考虑学习者知识结构、能力的变化,持续性地引导学习者进入下一个最适合的学习活动、为学习者找到与其当前知识和能力水平最匹配的学伴。

学习资源与服务智能推理引擎需要对学习过程中学习者表现出的个性特征(知识结构、学习情境、学习风格、学习情绪、与知识的学习交互等)进行语义建模,构建学习者的个性特征本体模型,设计多重影响要素的自适应聚合推理引擎,实现学习资源与服务的适应性推送。基本技术思路为:(1)研究学习过程中的认知规律,改进Felder-silverman、Kolb、Witkin等学习风格模型,LOCO、m-LOCO等学习情境模型和ARCS、FEASP等学习情绪模型,完善学习者模型;(2)在社会性知识空间构建基础之上,基于学习者个性特征本体模型,从学习风格、学习情境和学习情绪三个维度构建一个多重影响要素的适应性聚合模型。利用复合语义相似度计算技术和推理技术实现资源与学习者的适应性聚合的最优化,不仅聚合知识内容资源,而且聚合最合适的学习活动和人际资源;(3)最后利用可视化技术,结合认知地图,为学习者提供呈现符合自身个性特征的资源集合,包括个体认知地图、知识网络服务、资源学习顺序路径,为学习者提供有意义的资源组织形式和获取途径。

(七)区域教育决策分析

大数据可以使管理者从经验决策转型到数据驱动的决策,提高其教育管理的科学性^[8]。传统教育数据是基于抽样的局部数据,是在周期性、阶段性的评估中获得的,一定程度上不能完全客观真实地反映教育状况^[9]。同时,传统教育数据因为没

有合适的载体,只能随着时间流失,进一步加剧了数据的片面性问题。而随着数据采集与存储技术的发展,长期的积累可形成个体和群体大数据,通过对数据的横向和纵向的深度分析,教育管理机构能够及时准确地了解教育教学情况,预测并预警可能发生的不利事件,更加科学合理地做出教育决策。而且基于教育大数据的决策反馈过程更迅速甚至达到实时反馈,大数据具有易比较、交流、转化的特质,在此基础上做出教育决策,更符合当下信息时代要求及时响应的特点^[10]。

无论是宏观的教育制度与管理体制改革,还是微观的教学方法和管理方式的改革,都可以通过科学的数据分析寻找问题的症结所在,识别不同地区教育发展的独有规律,然后对症下药,实施改革^[11]。教育大数据能够聚焦于教与学的微观层面,将原本隐性的教育活动通过数据逐步清晰地描述出来,教育大数据也可以超越个体微观视野,从整体和宏观的视野来分析全局性问题,教育大数据还可以超越静态看问题的视角,从动态和发展的视角,去展现事物发展的过程,分析动态演化与发展过程中的问题。

由此可见,大数据思维帮助区域的教育管理摆脱传统的经验型被动决策,通过多元数据的科学分析、智能算法、依托区域教育质量地图,为教育管理者提供智慧的、科学的教育决策支持。

三、区域教育大数据的开放生态系统构建

区域教育大数据应用,只有平台和技术是不行的,关键要构建教育数据无缝流转的生态系统,数据要在不同用户角色、不同参与者之间、不同技术系统之间,在业务流程中无缝流转,在开展业务的过程中,既生成数据又使用数据,才能形成真正的大数据。靠行政推动的报送性的数据分析,永远无法形成真正意义上的教育大数据。

区域性教育大数据开放生态系统的构建,需要统一数据接入与交换格式规范,开放大数据分析服务,对内系统贯通,以单点登录模式统一用户、统一认证、统一结算,进一步实现业务系统之间数据与交互逻辑的衔接;对外开放大数据输入与输出、智能服务接入与输出的标准化接口,建立面向学习者、家长、教师、教育管理者和决策者的教育大数据可视化智能分析中心。通过各种智能终端应用、各种服务应用APP、各种智能教育装备系统采集数据,数据汇总后分析形成数据智能,通过API的方式,接入各个APP,提升服务应用的智能性,如下页图2所示。

- 统一用户、统一结算、汇聚大数据、汇聚资源、汇聚服务
- 开放服务接口、开放数据接口、开放用户服务市场

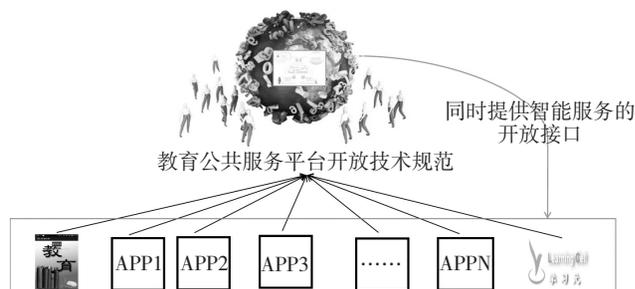


图2 区域教育大数据开放生态

区域教育大数据生态系统的建立，核心工作不是建设数据中心，而是促进系统之间的互联互通，是建立数据交换与贯通的技术规范与制度规范。

四、区域教育大数据的应用模型

本研究以智能教育大数据公共服务平台“智慧学伴”为例，介绍区域教育大数据应用模型，推进教育大数据应用要将学生放在数据应用的正中心，汇聚学生的学科素养、心理健康、身体发育、问题解决等方面的大数据，并在区域教育行政部门的牵头下前行，在学校和研训部门的多角色支撑下助力数据的深度应用，同时大数据也能更好服务家庭，促进家校协作，如图3所示。

区域教育大数据的应用可以使得家庭、学校、研训机构、教育行政部门之间，以学生数据为纽带，形成促进学生发展的紧密协同关系，能够有效提升区域教育质量。

(一)学生全面发展的数据汇聚

数据来源是区域大数据应用最重要的部分，在该应用模型中，由区域组织学生开展基于互联网的学科在线诊断，通过对学生学习数据的汇聚，对学生的知识和能力结合进行系统建模，建模数据类别主要包括学科能力测评、综合素质测评、双师在线服务、智慧学习伙伴、择校培优推荐、体质健康监测、问题解决提升、艺术素养挖掘等八个方面，数据汇聚的形式是作业、测试、问卷、体测、听课、线上行为、脑电波、问题解决等。通过后续的数据建模，可了解学生的能力素养和知识图谱，并通过大数据的分析为学生提供精准资源和在线教师及学伴服务，为个性化自主学习做好支撑。从区域提升的视角，学校的管理和学科教研实现定向改进，为更好的提升区域教育质量提供可行方案，下页图4所示为学科能力诊断大数据汇聚和应用的闭环流程。

(二)行政部门领航的大数据项目驱动

在大数据应用模型中，教育行政部门是“领路人”，在教育大数据落地中发挥着不可替代的作用，行政领航主要体现在面向主管领导、行政科室、信息化部门、教育督导处等层面的具体服务。

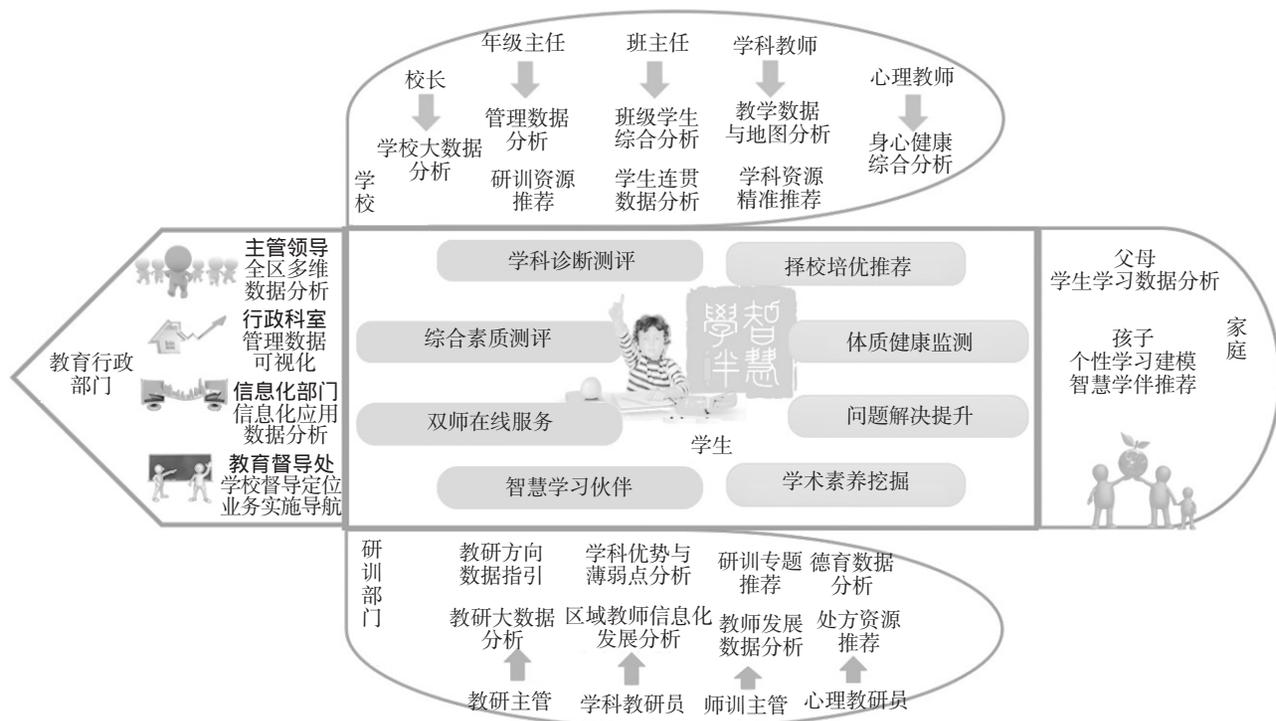


图3 区域教育大数据服务与应用模型



图4 学生大数据汇聚与应用的闭环服务

1. 全区多维度数据分析支撑行政领导宏观决策

大数据的战略意义在于“通过海量数据的交换、整合和分析，发展新的知识、创造新的价值，带来大知识、大科技、大利润和大发展”^[12]。行政主管领导一般指区域教育局长或教委主任等区域教育发展方向的掌舵者，大数据应用主要体现在宏观数据服务支撑决策上，包括教育质量地图、区域教育质量监测、区域资源配置、区域教育公共服务、教育发展规划制定等业务。全区多维大数据汇聚会形成区域教育可视化质量地图，质量地图包括各校教学质量热力图、学校升学率、学校师生比、教师组成结构等多维数据，帮助决策者定位区域不同层面质量问题，实现目标管理；全区多维数据分析可以支撑行政领导开展区域资源配置决策，如学生学业大数据反馈出来的区域师资不均问题，可通过区域层面的宏观调控，优化教师配置；依托区域大数据显现出来的教师群体信息分析、学生群体来源分析、学校发展建议分析等内容，可以支撑区域制定未来发展规划。

2. 管理数据可视化助力行政科室业务针对性推进

管理大数据可以服务区域中教科、小教科、人事科、教研室等不同业务直管部门，成为行政科室业务的加速器，实现数据支撑学籍异动、学段质量监控、教师能力提升培训等业务。由于社会人员的流动，高频的学籍调动带来了更多的转学、借读等业务的产生，学生的流动成为教育行政管理者的难题，智能平台和数据分析可以帮助行政管理者记录区域所有的学籍变动数据，行政科室可依据统计分析，及时制定出学籍变动的解决方案；历年学生中考大数据不仅揭示了学生发展的当时水平，更能通过多年连续大数据发现中学教育的发展走势，指导中学教育的发展改进；各个教师班级的教学质量数据，可以为师训部门组织教师专业培训提供参考。学生的能力发展数据、素养提升数据、师生关系数据的交叉分析，可以让管理者更清晰看到教师发展背后的问题，教师成长得到进一步的支持。

3. 信息化数据应用分析助推信息化部门建设

《教育信息化2.0行动计划》(以下简称“行动计划”)已于2018年由教育部正式发布，“行动计划”提出教育信息化从融合应用向创新发展演进，全面提升师生信息素养。将信息技术和智能技术深度融合教育全过程^[13]。可见，信息化部门不再局限于建网络、买设备，而是能借助大数据更有效的开展教育环境建设和解决教育资源均衡。信息化部门要保证区域大数据应用的基本网络和设备条件，更要关注师生大数据产生后对其业务的支撑，师生在线学习行为大数据会生成区域、学校、班级等层面的信息化行为报告(如图5所示)，这可以大大提升信息化部门的管理效率，实现精准定位，有效跟进。教育资源均衡不仅体现在信息化软硬件的配备上，更要体现在直接获得的服务层次上，要及时通过数据，了解到不同师生获得服务的机会，并及时调整信息与服务资源。

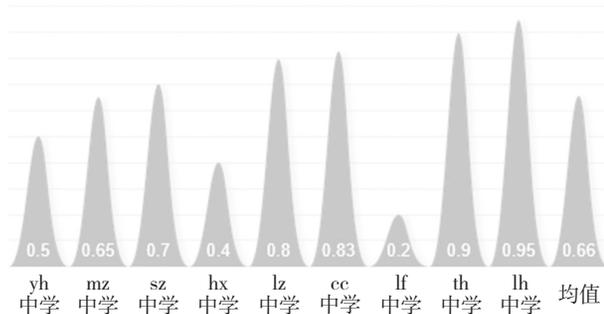


图5 某区各个学校语文教师月上线率

4. 大数据助力教育督导精准化

教育督导的主要职责包括常规教育督导，监督、检查和指导学校的教育领导和管理，对办学效益进行评估等，以前的督导主要体现在看某一个阶段的横截面，大数据提供了督导的新视野，可以看到发展和演化的历史过程，更好地开展评价导航，实现办学高效。还可以快速定位表象背后的关键问题，如透过大数据分析可发现某个区域某些学校的两级分化拉大，教育督导服务时可以直接督导该类问题的解决，直接定位到评价的关键环节，找到分化的原因，督导学校制定相应改进措施。

(三) 大数据推动学校管理和教学走向智慧

1. 大数据支撑学校行政智慧管理

智慧学伴大数据分析为校长和主任等教育行政人员提供了智慧管理的依据，可以依托数据进一步聚焦到学校发展共性个性问题、学生发展心理倾向、典型班级分析等业务上。大数据可以积累学校层面的教学数据、师生数据、班级数据，学校随时了解本校各学科在全区的相对位置，也可实现同

类学校之间的灵活对比分析(如图6所示),方便了解学校的共性问题,在共性的基础上,学校可以进一步透过数据分析问题产生的原因,聚焦到个体问题解决上,根据问题确定管理与改进措施。教育大数据能帮助学校识别典型班级,透过全校全年级的数据,分析发现不同年级或班级的学生能力素养、学习行为表现,依托数据发现异常班级,定向跟进,技术化解潜在管理问题。透过学校心理综合素养大数据,学校可以更好的开展德育教育,如心理健康、网络成瘾、师生关系优化等,既可以从学校整体视角改进,也可定位到不同个体,进行定向干预,最大限度促进每一位师生的身心健康发展。

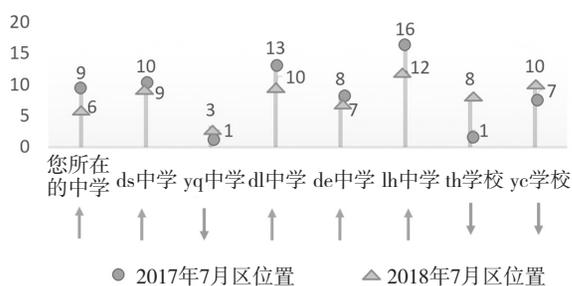


图6 同区各校学科大数据对比分析

2. 大数据支撑教师精准教学

大数据可直接助力教师精准教学,解决教学无据、学困生学困根源不明等问题。精准教学核心是通过学习数据分析,为每个学习者提供适合其最邻近发展区的学习服务与指导,通过学习行为数据的分析,教师可以精确了解学科教学是否达到目标、学生是否真正掌握知识或技能。“智慧学伴”可以通过对学生个体学习情况进行跟踪记录与分析之后,可以对学习情况建模,展现学生结构化知识地图,也可以展现全班整体教学薄弱点的教学地图,给予教师可视化的教学参考,让教学更有依据。右侧表所示是三个阶段利用大数据,帮助地理教师提炼出来的教学薄弱点,可供教师复习使用。学科学困生始终是教师教学颇为挑战的问题,智慧学伴可以帮助教师精准定位学困生的学习障碍点,让教师清晰了解学困形成的原因,从而制定解决问题的路径。

3. 大数据支撑班主任智能班级管理

面对一个班的几十名学生,每个学生个性特征、学习问题各异,班主任想做到因人而异的指导与干预,往往心有余而力不足,传统管理的低效更是让班主任知难而退。通过多侧面、多形态的数据收集,如生理数据、心理数据、学习过程数据等,进行多层次、长周期的数据分析与挖掘,可以帮助班主任完整了解每一个学生的显性或隐性的问题,

尤其是单亲等特殊家庭可能带来的学生成长问题,班主任借助数据分析可以精准问题定位、更加明晰管理重点。学生各有差异,所表现出来的学科优势也大相径庭,大数据支撑下,班主任可借助数据了解学生学科优势与学科短板,更好帮助学生进行中高考选科,实现最优化的组合。通过心理健康数据,班主任还可以了解学生家庭环境、人格障碍等引发学困的原因,从而更好的协同家庭,共同解决好学困问题。

地理学科薄弱教学点大数据自动汇聚表

地理学科	得分率		
	上学期末	下学期期中	下学期期末
核心概念			
地图	67%		
地球	72%		
中国水系与水文	59%		
中国自然灾害	57%		
中国天气与气候	63%	43%	90%
中国的地形	61%	67%	33%
中国人口	62%		56%
中国区域位置与分布	47%	55%	57%
资源与可持续发展		59%	78%
中国文化与旅游		79%	92%
中国经济		55%	58%
中国与差异与区际联系		63%	73%

注: 表示薄弱知识点 表示优势知识点

(四) 研训部门转型的大数据研修支撑

1. 大数据支撑研训主管决策教研的方向

在传统的研修形式下,研训主管了解培训成效和教师专业成长情况的渠道是下属工作汇报或专题交流,导致工作效率较低,时间周期较长,信息也缺乏准确性。大数据可以让研修决策更加专业和高效,解决研修无依据、教研活动无思路等问题,促进区域教研转型。智慧学伴服务平台依托大数据,可以将各学科的区域性共性问题可视化,指示教研业务推进的方向。基于数据,可以根据各个学科校际能力差异、全区各个学科教师的发展水平不同而合理配置研训资源;通过汇聚区域层面多个学段的学业大数据,分析不同学段学科发展薄弱点,可提升教研的精准性,如小学毕业后的科学学科数据与学生初中地理、物理、生物等学科数据的关联分析等,可以方便区域层面配置调整学段衔接的教研活动设计。

2. 大数据支撑教研员教研活动设计

教研员是教师教学的专业指导者,区域教学研究的组织者,教师专业发展的促进者,国家和地方改革政策的转化者,课程改革的推动者^[14]。在

区(县)教研工作中,教研员既要承担本区域的教学教研指导工作,又要承担本区域教师业务培训和教师职业能力提升工作^[15]。教研员日常的主要教研组织包括公开课研讨、专家培训、案例研磨等。智慧学伴大数据可以帮助教研员实现教研转型,带领教研员走出固有经验引领,主要表现在数据指引教研优势与薄弱点、区域教师信息化发展分析等。针对数据指引教研,教研员参考全区各校数据,基于连贯的多学科大数据发现学科教学的核心问题,如图7为学生提问数据汇聚后凸显出来的核心知识点问题,不同学校学科教师的发展短板亦可透过数据显现等。区域教师信息化发展分析是指教研员可以更好的管理区域本学科的教师,了解每位教师的在线情况,了解教师的专业投入度和学科资源应用大数据,并可根据区域教师群体的需求灵活调整教师教研计划,实现对教师业务能力的提升。



注:“其它”包含液体压强16、质量10、电能8、机械运动4、摩擦力3。

图7 学生提问大数据支撑关键内容教研与复习

(五)大数据支撑家校多维协同实现高效助学指导

要教育好孩子,家长必须和学校保持紧密联系,家校双方密切协作,家庭教育与学校教育紧密结合,这是实现育人目标的重要途径。在大数据助力区域发展系统中,家庭是大数据应用的重要场景,父母可以借助数据开展精准助学、基于数据实现更好的家校协同。大数据平台能够将家长和孩子连接起来,家长能够透过数据看到孩子发展的真实表现,借助大数据给孩子提供真正需要的助学指导,而不是盲目地报各种辅导班。如家长利用微信绑定孩子智慧学伴的教育账号,就可以了解到孩子的各种学业表现数据,可以精细知悉到孩子的学习优势和问题,为后续精准助学提供参照。

同样,孩子在家里也可以快速的感知自己的学习问题,并获得个性化的资源推荐服务、双师在线服务,还可以和大数据推荐的学习伙伴开展更有效的同伴学习,通过互联网和大数据,孩子可获得更多的学习小伙伴,延展了现实的学习圈。

综上所述,大数据应用模型从区域教育质量提

升的关键角色着力,提供了面向行政领导、校长、班主任、教研员和家长等角色的具体服务,解决了不同角色的业务关键问题,这是将数据资产最大化利用的抓手,也是大数据系统性解决区域问题的更重要武器,更是大数据能够落地的关键所在。

五、迈向智能教育新时代

区域教育大数据运行到一定时间,随着数据量的汇聚,到了一定的时候,会由量变迈向质变,形成智能化服务。其演化路径如图8所示,将分阶段达到“学习助手”“学习伙伴”和“学习导师”三个智能服务层次。

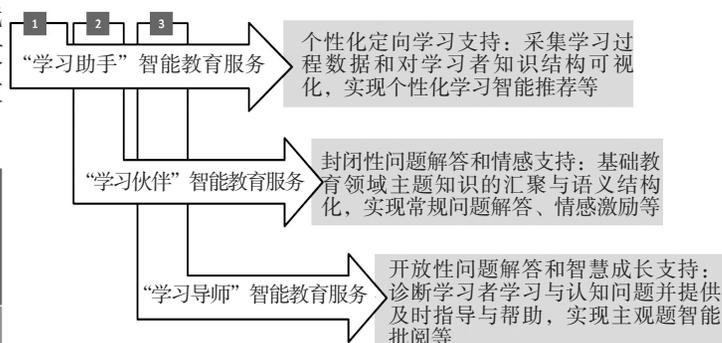


图8 智能教育服务的三个发展阶段

(一)“学习助手”级智能教育服务

第一个阶段,将使得教育公共服务平台达到学习者知识结构可视化及个性化学习智能推荐层次服务,能够担当家长、老师和学生的学习助手。具有自然全面采集学习过程行为投入数据、认知投入数据、情感投入数据,科学进行学习投入水平计算与评估,系统仿真学习者学科知识结构,自动生成面向学习者个人以及教师的可视化学习知识地图,帮助学习者找到与学习水平相适应的学习资源、学习活动序列、学习专家和学习同伴,帮助教师评估班级整体学习水平等特性。

(二)“学习同伴”级智能教育服务

第二个阶段,将使得教育公共服务平台仿真学习者的知识与认知模型,科学评估学习者学科认知水平,达到与学习者相当智能服务水平,可以在陪伴学习者过程中,进行封闭域问题和简答题的解答,进行情感激励与陪伴等,能够担当学生的学习同伴。具有半自动语义标注和结构化基础教育领域学科知识、学习资源,自动聚合同类知识主题和学科资源,智能分析与解答特定封闭域学科问题,客观分析与表征在线学习过程中的情绪状态,在线模拟学习计划编排等功能,为学习者提供常规问题解答、情感支持、学习伙伴等服务。

(三) “学习导师”级智能教育服务

第三个阶段,将使得教育公共服务平台能够诊断学习者学习与认知问题,及时予以指导与帮助,可以进行主观题的智能批阅,达到能够指导学生的相当教师智能服务水平,可以进行自主深度学习,实现进行开放域问题的解答,能够担当学生的学习导师。具有全面开放吸收学习过程中生成的智慧,自行积累、沉淀、重构与转化,主动学习外部知识,扩充内部知识结构体系,自动建立新旧知识多学科领域关联,智能分析与解答开放域学科问题,为学习者提供疑难解答、知识引领等特征。

参考文献:

- [1][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.
- [2] 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.
- [3] 万海鹏,李威等.大规模开放课程的知识地图分析——以学习元平台为例[J].中国电化教育,2015,(5):30-39.
- [4] ADL. xAPI Overview[DB/OL]. <https://www.adlnet.gov/research/performance-tracking-analysis/experience-api/>, 2018-08-21.
- [5] 万海鹏,余胜泉.基于学习元平台的学习认知地图构建[J].电化教育研究,2017,38(9):83-88.
- [7] 余胜泉,杨现民等.泛在学习环境中的学习资源设计与共享——

“学习元”的理念与结构[J].开放教育研究,2009,(1):47-53.

- [8] 余胜泉,王阿习.“互联网+教育”的变革路径[J].中国电化教育,2016,(10):1-9.
- [9] 刘雍潜,杨现民.大数据时代区域教育均衡发展新思路[J].电化教育研究,2014,(5):11-14.
- [10] 陈霜叶,孟浏今等.大数据时代的教育政策证据:以证据为本理念对中国教育治理现代化与决策科学化的启示[J].全球教育展望,2014,43(2):121-128.
- [11] 王丽珍,刘佳星.教育大数据对网络教研的影响[J].中国电化教育,2016,(11):51-55.
- [12] 涂子沛.大数据:正在到来的数据革命,以及它如何改变政府、商业与我们的生活[M].桂林:广西师范大学出版社,2012.
- [13] 教育部.教育部关于印发《教育信息化2.0行动计划》的通知[DB/OL]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html,2018-06-30.
- [14] 刘月霞.中国教研的价值与使命[J].基础教育论坛,2016,(26):15-16.
- [15] 黄生英.教研员:担当起自己的多重角色[J].湖南教育(D版),2016,(10):22-23.

作者简介:

余胜泉:教授,博士生导师,研究方向为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(yusq@bnu.edu.cn)。

李晓庆:硕士,研究方向为信息技术与课程深层次整合、技术支持教与学、大数据助力区域教育质量改进(lixiaqing8507@bnu.edu.cn)。

The Study on Architecture and Application Model about Regional Education Big Data

Yu Shengquan, Li Xiaoqing

(Beijing Advanced Innovation Center for Future Education, Beijing 100875)

Abstract: Big data Education, as an opportunity for future development of education in regions and a method for quality improvement, has immeasurable development potential in the field of intelligent education. From the perspective of the relationship between big data and region education development, this paper analyzes the technical architecture, which contains modal data collection of big data in regional education, personalized knowledge model building about learners, subject knowledge map construction, data mining and analysis, resource semantic markup and gathering, personalized intelligent recommendation engines, regional education decision analysis, etc., and put forward the construction of open ecosystem that regional big data can be transferred seamlessly. Meanwhile, Taking the intelligent education big data public service platform “smart learning partner” as an example, this study proposed the big data application model in region education, and make outlook about the three intelligent service stages about big data of region education in the future development: “learning assistant” with personalized directional learning support, “learning partner” with closed questions and emotional support, “learning mentor” with open questions and intelligent growth support. This study comprehensively discussed big data of region education from three aspects of theoretical construction, practical application and future prospect, expecting to provide theoretical reference model and practical guidance for the application of big data in region education.

Keywords: Big Data of Regional Education; Improve the Quality of Regional Education; Application of Big Data in Education; Ecosystem of Big Data in Education

收稿日期: 2018年10月20日

责任编辑: 邢西深 赵云建