

教育大数据驱动学业诊断的思考与探索

孙佳¹,李维²,王菲菲¹,魏旭¹,盛叶娟¹

(1.北京爱云校教育科技研究院,北京 100020;

2.深圳市福田区红岭中学(红岭教育集团),广东 深圳 518000)

摘要:本文基于教育大数据建设和应用的视角,探讨教育大数据驱动学业诊断发展的相关问题。文章认为,教育大数据驱动了学业诊断的精准化、普及化和常态化发展,指出教育大数据要实现对学业诊断发展的驱动,需要解决好三方面问题:数据采集的全面性,包括学习结果数据、学习过程数据和学生发展数据;数据分析方法的多样性,深入分析结构化数据,并突破非结构化数据分析的难点;数据应用的便捷性和针对性,满足不同教育主体对学业诊断的数据分析需求。本文还基于实践,介绍了教育大数据驱动学业诊断在促进业态升级、学校教育发展和学生成长方面已取得的现实进展,做到了减负增效,促进了学生全面而个性化的发展。

关键词:教育大数据;学业诊断;学习分析

中图分类号:G40-058.1

文献标志码:A

文章编号:1673-8454(2019)15-0048-05

随着基础教育课程改革的全面深化和新高考改革的推进,发展学生核心素养和实现“立德树人”育人目标备受全社会关注,促使教育实践领域愈加重视“学业诊断”。近年来教育大数据的迅猛发展,为学业诊断注入新思路、新动力、新引擎。本文探讨了教育大数据驱动学业诊断的现实意义,提出其中需要把握的关键问题,并且介绍了相关实践进展。

一、教育大数据为学业诊断发展提供机遇与实践推广的可能性

1.教育大数据驱动学业诊断实践走向精准化

学业诊断是医学领域的“舶来品”。由于教育教学和医学都面临复杂的、不确定的情境,有学者提出教学向医学学习,为学生开出“化验单”、“报告单”和“诊断书”。传统学业诊断如同大多数教育决策判断一样,由于数据范围有限、分析方法有限,更多依靠经验判断,精准程度仍有很大提升空间。^[1]

教育大数据丰富了学业诊断的数据类型,为其提供了更多的科学分析方法的应用空间,能从多角色的角度更客观地描述实际情况:通过海量的学习数据,精准分析每位学习者的知识能力结构、个性倾向、思维特征、学习路径和学科素养发展状况。^[2]学业诊断突破原有技术条件限制,形成新样态:基于育人目标和核心素养,采集全样本、全过程数据,运用多种分析方法,从认知、行为、情感三个维度分析学习结果、学习过程和成长发展相关因素,全面识别学生的学习现状与发展趋势、判断学生

的优势与不足,从而为后续发挥优势和改善不足提出合理建议。以学年诊断为例,越来越多的学校不仅看重学生期末考试数据,还融合过程性评价数据,如笔记、日常作业、课堂表现、单元测试、专题测评、项目学习等,使学业诊断更加客观、连贯、精准。

2.教育大数据驱动学业诊断实践走向普及化

传统学业诊断方法的应用受到多种因素阻碍,比如理论模型比较复杂、技术实现过程困难,通常需要评估机构、教育测量学专家、心理测量学专家与学科专家协作。因此,以往学业诊断更多停留在研究领域或小规模应用,学校实践领域的专业化学业诊断往往由学校聘请专家对重点年级或班级进行诊断。

随着教育大数据不断发展,“互联网+教育”的平台和工具不断完善。通过系统建构一定的参数和模型,学业诊断的门槛降低,教师利用平台和工具进行高质量诊断成为可能。学业诊断的主体逐渐从专家拓展到学校、教师和学生个体。越来越多的学校借助诊断系统,实现教师主导诊断、学生自我诊断。以“好分数”测评分析系统为例,已经有3万所学校、150万教师、3000万学生利用教育大数据实现学情诊断。同时突破校际的统考、联考活动也越来越频繁,既有区域联考、盟校联考,也有跨省市数百所学校的诊断性联考。

3.教育大数据驱动学业诊断实践走向常态化

在传统教学模式中,教师很难实现面向学生个体的动态学业诊断。在课堂上,教师通过观察、互动,课后通

过作业和试卷,或通过与学生对话进行学情诊断,但是这些往往都属于经验性的、瞬时性的信息收集、点状诊断,很难持续收集学生的信息进而形成动态诊断。

教育大数据正逐步实现对学习全程实时的数据采集、诊断评估和指导反馈。随着智慧校园的建设,信息化平台在校内大范围、高频率的应用,学生信息的收集与记录做到了实时,为学业诊断提供了连续的证据,使常态化的诊断得以实现。比如,学生在在线课堂的学习路径会被自动记录。以云课堂为例,系统能够自动记录学生选择的学习资源类型、观看微课的时长、浏览学案的完成度、课后作业准确率、小组协作的成果等,通过对学生学习结果和过程的常态化追踪、分析,识别学生的学业水平、学习行为与能力发展。

二、教育大数据驱动学业诊断需要解决好三方面问题

1. 数据采集问题

确保大数据的可用性必须从源头抓起。^[3]数据采集决定学业诊断的全面性与科学性。因此,用于学业诊断的教育大数据应当是子集丰富的数据集。然而,在当前教育实践中,数据采集过于关注学生考试测评等结果数据,学习过程数据缺失,学生个性特征、发展目标等数据不足。因此,学业诊断的证据不充分,影响诊断结果的准确性,对学生发展的预测力不足,需要对数据采集、分析的范围进行限定,详见图1。



图1 应用于学业诊断的教育大数据组成

全面的学业数据包括三类：第一类是学习结果数据。学习结果数据指的是终结性评价数据,来源于期中考试、期末考试和大型测评。数据集包括期中、期末考试成绩和统考、联考成绩。第二类是学习过程数据。学习过程数据指的是形成性评价数据,来源于传统课堂、在线课堂和综合实践。数据集包括考勤数据、课堂交流数据(口头回答、口头汇报,在线文字交流)、阶段测评数据(随堂作业、课后作业、小测验等)和操作表演类数据(如

实验、小制作、角色扮演等)。^[4]第三类是学生发展数据。学生发展数据指的是学习规划及相关因素的数据,来源于排选课、生涯规划和学校教务管理。数据集包括选选课数据、生涯测评数据和成长档案数据。

2. 数据分析问题

在学校学业诊断中,教师 and 教学管理者大多使用统计分析法,针对学业成绩、知识点、能力结构等结构化数据,进行初步的描述性分析。同时,在传统教学中,师生的言语数据、行为数据丰富,但对这些非结构化数据的分析能力不足,多数教师无法从中深度解读出学业信息。随着大数据技术发展,更多数据分析方法进入诊断领域。研究指出,未来学校学习中,将有超过80%的数据是非结构化数据。^{[5][6]}因此,学业诊断应综合应用各种分析方法,深入分析结构化数据,并突破非结构化数据分析的难点,对收集到的学习结果数据、学习过程数据、学习发展数据进行描述性和预测性分析。

因此,教育大数据驱动的学业诊断可以借助测评分析系统,综合分析全过程学习数据。使用统计分析法,整合经典测量理论、项目反应理论和认知诊断理论优势,分析学习结果和学习过程中的结构化数据。借助研究者的实践经验,采用先进的分析方法破解非结构化数据分析难的问题,挖掘学习过程数据和学生发展数据的教育价值。比如利用文本分析法、话语分析法,提取师生交流、生生交流中隐藏的关键信息,实时跟进学习者的意见和想法。利用社会网络分析法,通过在线行为的分析,发现学生社会网络特征变化的特点,以及学生社会网络位置与学习成效之间的相关性。^[7]

3. 数据应用问题

在学业诊断过程中,采集、存储、分析数据,目的是应用数据、读懂学生、服务教学提升。因此要重视数据应用的便捷性和面向不同教育主体的针对性。只有把数据应用和教学深度融合,才能提升干预的有效性。然而,在学校教育实践中,有些学校学业诊断系统形同虚设,或者教学应用与数据采集、分析脱节;有些学校混用多套测评分析系统,无法保证数据全面且持续;有的诊断系统采用数值化、表格化、堆积式等方式呈现诊断结果,可读性较差。^[8]只有将学业诊断建议转化成教学提升行动,学习反馈才更及时,学习干预也更精准。

因此,要想利用教育大数据驱动学业诊断,不仅要改善诊断结果呈现的方式,借助图像、图形、多元素关联图表等,使各个使用者一目了然,快速理解诊断信息。更要重视诊断结果的针对性,满足学生、教师、教学管理者差异化的需求,使学业诊断服务学生自我反思、自主学

习、自主规划;服务教师干预学习、教学反思、教学改进;服务教学管理者完善教学支持、提升决策质量、改善评价方式。

三、教育大数据驱动学业诊断的实践探索

1. 行业实践探索

随着大数据发展,行业从业者积极探索,与高校、中小学、科研机构深度合作,全面理解、深度挖掘业务需求,从教育教学视角思考教育大数据产品的体系框架与流程设计。^[9]学业诊断在单一场景的应用和多场景综合应用方面都取得了进展,促进了业态升级。比如在线口语考场,通过人机对话,诊断学生的语言能力;阅卷与测评分析系统,采集分析学生考试数据,诊断学生的知识结构和能力发展情况。也有平台通过全过程数据采集、分析,实现动态化、个性化的综合性学习分析,引导学生开展个性化学习。

以爱云校为例,研发中心与北京师范大学脑与学习科学研究中心、北京十一学校、烟台市教科院、青岛黄岛区及全国多所中学合作,开发“全过程学业数据采集与分析诊断云平台”。云平台整合智能组卷系统、测评分析系统、过程性评价系统、在线课堂系统、排选课系统、生涯规划系统和教务管理系统,追踪和分析学生学习结果、学习过程和学习发展数据。利用全过程数据,构建动态、完整、个性化的学生画像。并根据学生个性特征差异,匹配适宜的知识路径,为教师、学生、管理者提供数据参考和决策建议。

借助“全过程学业数据采集与分析诊断云平台”支持,数据采集和分析流程如图2所示。首先,通过阅卷系统自动生成考试数据分析,由此了解学生学科知识点掌握情况、能力发展情况、学科均衡性以及成绩动态变化等学业信息。然后,通过过程性评价、在线课堂,掌握学生学习行为表现(学习时长、使用资源类型、学习进度及结果等),作业完成度(正答率、薄弱知识点等)以及师生、生生交流情况等。最后,借助排选课系统、生涯规划系统等,结合学生选选课结果、生涯测评结果和成长档案,对每一位学生的优势和不足做出综合性的反馈。



图2 云平台支持学业诊断系统示意图

2. 学校教育的应用实践

在学校教育中,基于教育大数据的学业诊断已经渗透到教学全过程。针对测评数据,系统越来越自动化、智能化,科学性和便捷性也显著提升。比如,学校借助“好分数”测评分析系统,实现了数据采集智能化、数据分析自动化。(见图3)一方面,学生手写内容的智能识别已经取得突破性进展,比如考试测评、日常作业、学生作品中的部分手写内容已经能通过机器自动识别,完整保留了学生的答题痕迹,也在不改变教师使用习惯的前提下,减轻教师手动处理的工作量。另一方面,依托知识图谱,系统智能识别测评内容对应知识点和能力维度,自动生成学科整体情况、班级整体情况、临界生分析、优劣势学科、命题质量、历史对比分析等多维度分析报告,帮助教师快速了解教学中的关键问题。

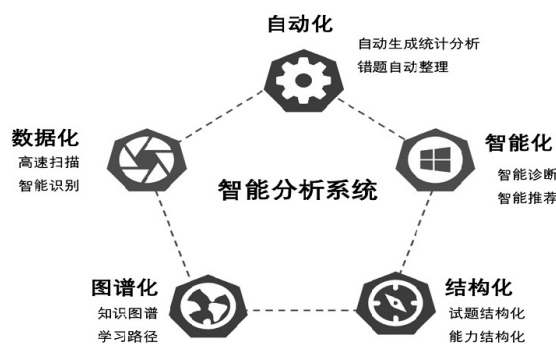


图3 好分数智能分析系统

与此同时,针对学习过程数据不易追踪、分析的状况,学校实践也有突破。比如S学校借助在线课堂系统分析学生线上学习过程。采用文本分析和社会网络分析相结合的方法,分析课堂讨论区学习数据;基于教师提供的多个开放性问题,通过分析学生问题选择的偏好、回帖次数、回复内容的思考深度,探究学生的学习态度、认知水平、思维方式。通过查看师生和生生交互频次、点赞次数等分析学生课堂参与、小组合作贡献度等多项重要信息,识别学生的合作与交流素养等。再如,学校借助过程性评价系统分析学生在传统课堂的学习过程。L学校按照出勤和行为规范、合作学习、课堂表现、家庭作业和测验,设定多维度指标体系,并且进行赋分,如图4所示。而N学校则关注学业发展、互动合作、公民素养,由此将笔记及任务完成情况、学科参与(师生互动、生生互动)、考勤纪律等设定为校本指标。

3. 学生发展的应用实践

教育大数据驱动的学业诊断,让学生参与到诊断过程中,更清晰地认识自己、唤醒自己、自主规划、发挥学习主体的作用。通过查看全过程学业分析报告,学生能



图4 L学校使用过程性学业评价系统示意

更准确地获取和分析自身的学习进度、学习习惯、学习偏好以及个性特征等信息,突破以往以成绩作为单一维度自我评价的模式,引导关注成长历程中的自我对比、正视不足、肯定进步、发现优势与潜能、激发成长内动力、整合促进自身成长的资源。

在教育实践中,已经出现利用大数据改进学业诊断以助力学生全面而个性化发展的实例。比如,W同学借助“好分数”学业诊断系统,把学习现状、发展趋势、个人特质与升学目标结合起来。(见图5、图6、图7)借助系统,W同学了解自己的薄弱知识点、学业成绩动态变化和学科优劣势。利用系统自动推送的匹配练习,结合教师的差异化指导,查缺补漏、精准提升。借助过程性评价和在线课堂相关数据,看见自己学习方式偏好、识别交流模式,透过同伴和教师的评价,加深自我认知,改善人际交往模式。借助生涯规划系统,评估兴趣、性格、能力与学业的匹配程度,确定选课选科与成长规划,唤醒自主意识、提升学习的积极性和主动性。

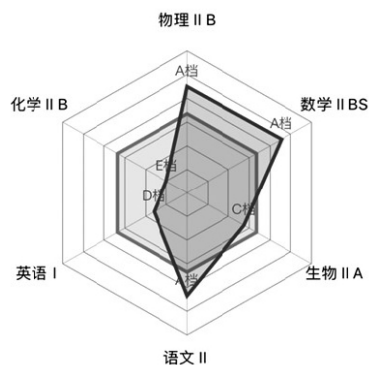


图5 W同学学科优劣势分析图

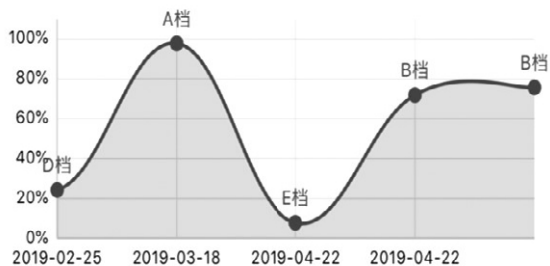
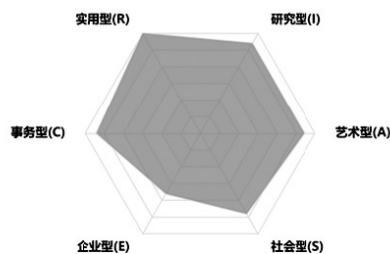
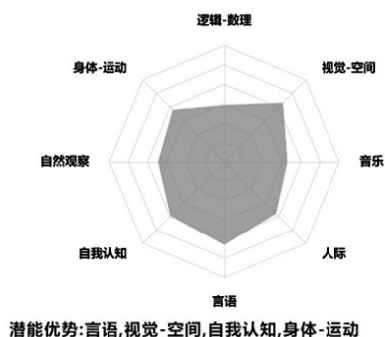


图6 W同学成绩趋势图



兴趣类型:RIC



潜能优势:言语,视觉-空间,自我认知,身体-运动

图7 W同学个人特质示意图

四、结束语

在数据驱动教学的背景下,学业诊断在减负增效、促进学生全面而个性化发展方面发挥越来越重要的作用。基于教育大数据的学业诊断正在建构以“学习者为中心”的减负增效新模式:全过程、全方位了解学生的学习进程,精准识别优势和不足,及时发现显性和隐性问题;通过教学干预帮助学生调整学习状态,规划合理学习路径,助力学生“自主减负”、“科学增效”。由于教育过程的复杂性和多样性,基于教育大数据的学业诊断仍然面临许多挑战。然而,随着5G技术、物联网、人工智能的进一步发展以及教学实践的不断推进,学业诊断的数据采集及分析方法将会进一步丰富,数据应用会继续揭开学习过程的黑匣子、为学生发展创造新的教育价值。教育大数据在赋能学业诊断的同时,也将会被学业诊断重新塑造。

参考文献:

- [1]刘凤娟.大数据的教育应用研究综述[J].现代教育技术,2014(8):13-19.
- [2]余胜泉.大数据时代的未来教育[J].中国民族教育,2017(Z1):8-11.
- [3]杨现民,唐斯斯,李冀红.教育大数据的技术体系框架与发展趋势——“教育大数据研究与实践专栏”之整体框架篇[J].现代教育技术,2016(1):5-12.
- [4]崔允漷.学校课程实施过程质量评估[M].上海:华东师范大学出版社,2017.8:206.

Python 数据挖掘在高校人才引进中的应用 *

陈 镭¹, 张凡龙²

(1. 南京审计大学 人力资源部, 江苏 南京 211815;

2. 南京审计大学 信息工程学院, 江苏 南京 211815)

摘 要:随着大数据时代的来临、互联网技术的快速发展,如何从海量数据中挖掘出潜在的、有价值的信息,成为当下热门的研究和实践方向之一。Python 因其强大的科学计算能力成为数据挖掘领域不可或缺的工具。基于 Python 语言对高校人事、科研及互联网数据进行挖掘和分析,构建人才发现系统,为高校引进高水平人才提供决策支持。

关键词:Python; Scrapy 网络爬虫; 数据挖掘; Sklearn; 人才引进

中图分类号: TP311

文献标志码: A

文章编号: 1673-8454(2019)15-0052-03

一、引言

大数据时代,数据的规模在不断扩大,数据呈现出体量大、多样性、价值密度变低、关系复杂的特点。从一般的信息系统中获取具有决策价值的信息变得越来越困难。如何做好高校人才队伍长期发展规划,满足高校快速发展当中对高水平人才的需要,人事管理部门迫切需要拥有智慧的信息系统,对目前掌握的大量人事、科研、教学信息进行分析、挖掘,对学校引进人才进行科学规划与评测,对在校人员进行科学培养与管理提供决策支持。

数据挖掘技术是一门涉及数理统计、人工智能以及知识工程等领域的新兴交叉学科,包含一套完整的过程。^[1]该过程首先从数据库、数据仓库或大数据集中抽取潜在有价值的知识、模型或规则,然后使用这些知识做出决策。换一种说法,数据挖掘技术是通过多种复杂算法,在大量未加工数据中发现数据间潜在关系以及有效知识。近年来,数据挖掘技术在金融业、零售业、医疗和电信等多种行业都取得了显著成果。

Python 语言是当今最热门的程序设计语言之一。它是一种面向对象的语言,拥有较好的交互性、解释性。

* 基金项目:本文系国家自然科学基金青年项目“面向鲁棒特征提取的矩阵秩极小化问题研究”(61603192)。

[5] Steve Kearns. Six Reasons Businesses Need to Pay Attention to Unstructured Data [EB/OL]. <http://www.itbusinessedge.com/slideshows/six-reasons-businesses-need-to-pay-attention-to-unstructured-data-05.html>.

[6] Yuejin Xu, Noah Reynolds. Using Text Mining Techniques to Analyze Students' Written Responses to A Teacher Leadership Dilemma [A]. Proceedings of the 2011 4th IEEE International Conference on Computer Science and Information Technology [C]. New York: ACM

[7] 石月凤, 刘三妍, 刘智, 韩继辉, 彭晔. 基于社会网络分析的在线学习行为分析实证研究[J]. 中国教育信息化, 2019(1):5-10.

[8] 刘三妍, 周东波, 李浩, 孙建文, 于杰. 基于地图的教育大数据可视分析方法探讨[J]. 电化教育研究, 2018(7):49-56.

[9] 杨现民, 田雪松. 中国基础教育大数据 2016-2017: 走

向数据驱动的精准教学[M]. 北京: 科学出版社, 2018.3:224.

[10] 罗照盛. 认知诊断评价理论基础[M]. 北京: 北京师范大学出版社, 2019.2:3.

[11] 付志慧. 多维项目反应模型应用理论[M]. 北京: 科学出版社, 2017.3:3.

[12] 张琪编著. 学习分析技术与方法[M]. 北京: 科学出版社, 2018.12:6-7.

[13] 刘月霞, 郭华主编. 深度学习——走向核心素养[M]. 北京: 教育科学出版社, 2018.11:6-7.

[14] 刘三妍, 杨宗凯. 量化学习——数据驱动下的学习行为分析[M]. 北京: 科学出版社, 2016.6:85.

[15] 祝智庭, 孙妍妍, 彭红超. 解读教育大数据的文化意蕴[J]. 电化教育研究, 2017(1):29.

[16] 孔兴隆. 运用学业质量检测数据, 诊断与改进学生的学习[J]. 考试研究, 2016(3):63.

(编辑: 王天鹏)