

教育大数据流通现状及对策研究

李成¹, 张人杰²

(1.长沙民政职业技术学院, 湖南长沙 410004; 2.湖南邮电职业技术学院, 湖南长沙 410015)

【摘要】教育大数据是人才培养过程中产生的数据集合,在教育教学、管理服务、决策支持中发挥着重要作用。结合我国高校在开展教育大数据采集、治理、存储、共享等方面的现状,探究了现有问题的主要成因,并通过借鉴国内典型案例,总结归纳了教育大数据流通实施的对策,对于深化教育教学改革、提升治理水平具有借鉴价值。

【关键词】教育大数据; 流通现状; 信息化治理

【doi:10.3969/j.issn.2095-7661.2022.03.013】

【中图分类号】G434

【文献标识码】A

【文章编号】2095-7661(2022)03-0052-03

Research on Current Situation and Countermeasures of Education Big Data Circulation

LI Cheng¹, ZHANG Ren-jie²

(1. Changsha Social Work College, Changsha, Hunan, China 410004;

2. Hunan Post and Telecommunication College, Changsha, Hunan, China 410015)

Abstract: Education big data is a collection of data generated in the process of talent training, and plays an important role in education, management services and decision support. In combination with the current situation of universities in China in the collection, management, storage and sharing of education big data, this paper deeply explores the main causes of its problems, and summarizes the countermeasures for the implementation of education big data circulation by referring to domestic typical cases, which has important reference value for deepening education and teaching reform and improving governance level.

Keywords: education big data; current situation of circulation; governance by information technology

在移动互联网、云计算、物联网等技术高速发展的浪潮下,大数据正驱动着高校在教育教学、管理服务、资源共享等方面进行深入改革,也为构建现代化治理体系和提高现代化治理能力奠定了坚实基础。教育部2018年印发的《教育信息化2.0行动计划》^[1]明确提出,充分利用云计算、大数据、人工智能等新技术,提升教育教学现代化水平,文件对于大数据在高校的应用提出了更高的要求。但纵观高校推动教育大数据建设与应用的现状,仍面临着数据来源不唯一、数据标准不统一、数据准确性得不到保障、缺乏有效的数据交换与共享机

制等数据流通方面的问题。要破除以上困境,激发大数据在高校人才培养中的更大潜能,还需进一步探讨解决数据来源统一、异构数据融合、常态化数据治理、智能化决策支持等方面的问题^[2-3]。

1 教育大数据的概念界定及特征

“大”到什么程度的数据可以称为大数据?网络上对大数据的特征归纳为5V,具体为Volume(大量)、Velocity(高速)、Variety(多样)、Value(低价值密度)、Veracity(真实性)。

1.1 Volume(大量)

量大如何界定?以高校学生成绩数据为例,学

【收稿日期】2022-05-27

【作者简介】李成(1985—),男,湖南长沙人,长沙民政职业技术学院实验师,硕士,研究方向:软件工程、数据治理。

【基金项目】2020年湖南省教育厅科学研究项目“基于区块链的教育大数据流通可信机制及关键技术研究”(项目编号:20B045);2020年教育部产学研创新基金项目“大数据背景下网络安全轻量态势感知平台的设计研究”(项目编号:2020ITA07026);2021年湖南省教育科学研究工作者协会科研课题“高职在线教学资源库建设研究”(课题编号:XJKX21A056)。

校通常记录每学期学生期末成绩,也就是说一年记录两次。以在校生2万、平均每学期开课10门为例,加上可能会有学生出现补考、重修等情况,一年记录的成绩数据约为40万条,10年后收集的数据就约为400万条,这些数据量足够供学校分析学生的学习情况存储相关行为。这些数据的量与商业系统的数据体量对比不大,但对于高校研究学生、课程的成绩与相关过程来说,这些数据已足够可以成为样本,所以称其为大数据也不为过。

1.2 Velocity(高速)

高速是指数据在单位时间内具有快速的增长频率与更新速度,如国内某电子商务系统,在高峰期间每秒的商品浏览记录可以达数亿条,记录着客户来源地区、浏览的时间、商品ID等信息,更新频率极高,完全符合大数据高速的特征,对于大数据分析用户偏好、商品需求具有非常重要的支撑作用。在高校这种情况不会存在,高校常见更新速度快的数据如校园一卡通消费、门禁、在线教学行为等,可能在高峰期也就每秒数万条。但对于高校利用大数据分析学生消费习惯、贫困生资助、人员流动管控等,其更新频率与速度可以满足作为数据样本需求。

1.3 Variety(多样)

多样一般可以从两个方面进行反映:一是同一条数据中的数据项多,比如学生基本信息包括性别、身高、体重、出生日期、出生地、民族、婚姻状态、政治面貌、健康状态等方面的特征;二是数据来源种类多,比如分析学生的学业进度,可能包括课程的成绩、第二课堂的学习情况、日常行为的表现等,其数据来源分布在不同种类的信息系统。

1.4 Value(低价值密度)

低价值密度是指大数据的价值密度相对较低。数据的价值密度和数据的规模呈反相关,数据的规模越大,数据的价值密度越低。大数据最大的价值即在于从大量低价值密度数据中挖掘出对分析和预测等有价值的信息,比如网络设备、服务器的日志文件,包含着设备本身运行状态信息、用户连接信息等,其数据产生频率高、体量大,但有价值的信息较少。

1.5 Veracity(真实性)

数据的真实性是指在数据的生命周期内,其完整、一致、准确的程度。对于高校来讲,大数据的来源广泛、采集流程复杂、人为影响因素较多,导致数据的真实性很难得到高度保证,这也是大数据技术重点需要解决的问题之一。比如学生的银

行卡信息,因涉及消费、资助等业务,导致数据来源不唯一,数据准确性不高。

2 教育大数据流通现状及存在的问题

当前,高校在智慧校园硬、软件条件上逐渐成熟,通过将人才培养、学术研究、管理服务与校园生活等应用系统进行整合,有效地提高了校园数字化水平与管理效率,也积累了大量的数据资产,但是信息化建设是有阶段性的,由于缺乏统一规划、各部门独立建设,信息孤岛严重。

2.1 数据来源不唯一,数据重复采集现象严重

一方面因为在传统信息化管理情况下,各职能部门之间协同、沟通较少,管理职能存在重叠,导致数据采集的界限不清楚。二是常用数据在各个职能部门管理的信息系统中都进行了采集,例如学生的照片信息,同一个数据在学籍系统、学工系统、校园一卡通系统都进行了采集,但是数据各不相同,那么以谁的数据为准。

2.2 标准不统一,多源、异构数据统一存储困难

一方面是高校在十二五、十三五期间,已经历多轮信息化基础设施、信息系统的建设与升级,建设时基本是以解决业务需求、痛点为出发点,各系统的数据存储的标准不一,比如师生性别,人事系统以“F”“M”进行男、女性别的存储,但在教务系统却以“1”“0”进行存储,导致数据融合困难。另一方面,大数据的来源不再局限于结构化数据,非结构化与半结构化数据也是其重要组成部分,传统的关系型数据库无法对异构数据进行统一存储、计算。

2.3 缺乏数据共享机制,信息系统孤岛普遍存在

一方面缺乏统筹规划,对信息系统建设的前瞻性、可行性研究不够深入,导致各职能部门在系统建设时各自为政,数据集成困难。另一方面信息化资金投入不够,难以按照预设目标进行全面建设,分期建设导致各个子系统存在代差,数据库之间相互独立,信息孤岛严重。最后,由于部门系统功能受上级主管部门监管,出于数据的专业性、保密性等方面考虑,难以进行数据开放与共享。

2.4 缺乏支持数据平台,监控、分析和预警不及时

尽管移动终端系统的不断发展为信息采集带来了极大的便利,但采集后的数据仅仅停留在查询阶段,缺乏一个支持决策的大数据分析平台,不能及时对人才培养状态数据进行有效监控、分析和预警。

2.5 缺乏可靠的安全机制,数据流通安全性差

2019年教育部印发《教育行业密码应用与创新实施发展方案》,明确提出以“网络实名管理、重

要数据加密、个人隐私保护”为应用目标,对学生、教师、学校等三大类基础数据产生、存储、使用、迁移、归档、销毁等业务环节进行全方位防护,完善系统身份认证、访问控制、权限管理、安全审计等保障机制^[3]。然而,教育大数据的采集源头多、受访问数据的主体集合构成复杂、流通场景和安全需求不确定性等特征,导致在流通过程中存在安全性差、防篡改性弱、不可溯源、个人隐私保护机制差、授权管理不可控等问题。

3 教育大数据流通机制与策略的研究

大数据时代与相关技术的出现,为教育大数据流通提供了新思路,利用其对分散在不同系统、不同设备上的异构数据进行有效整合与管理、深度挖掘与分析,进而发现数据所隐藏的价值,为高校治理能力提升、业务流程优化、辅助决策支持等提供支撑,进一步理解与发现师生的潜在需求,实现更加个性化、精准化的管理与服务^[4]。

3.1 构建数据标准与UC矩阵,实现数据来源与标准统一

围绕学校治理体系,全面梳理数据资源目录和数据溯源图谱,制定《信息化数据编码标准与规范》《标准代码集》,划清各部门数据管理界限,解决各职能部门数据共享没有标准可依的难题,实现数据源头统一、标准统一、数据统一,打好校园数据共建共享的基础。数据UC矩阵样例如表1所示,数据项标准样例如表2所示。

表1 数据UC矩阵样例表(U代表使用、C代表产生与更新)

数据子集单位	部门1	部门2	部门3
学校基本数据	C	U	U
办学条件数据		C	U
机构设置数据	C	U	

表2 数据项标准样例表

编号	字段名称	字段代码	类型	长度	主键	空值	可选	取值范围
1	学校标识码	XXBZM	N	10	Y	N		
2	学校名称	XXMC	N	60	N	N		

3.2 构建全量数据仓库,实现源头数据融合

在现有教学资源、教务管理、教学平台、人事管理、资产管理、财务管理等信息系统,各种设备设施,相关网站等的基础上,建立数据采集平台,对多源、异构数据进行同步采集、治理并形成数据仓库,在标准数据集的基础上建立应用,最终实现人、财、物等教育大数据的互通。

3.3 构建数据开放共享平台,实现信息化治理常态化、协同化

依托全量数据,制定数据隐私与开放制度,建立健全人才培养全量数据隐私、伦理安全规范与安全保护机制,通过数据发布、数据加密、访问控制等手段提供统一数据接口,建立各职能部门数据开放共享、互联互通机制,解决学校业务系统独立、数据孤岛等问题,为常态化、协同化、信息化治理工作提供数据保障。

3.4 构建校园大数据分析平台,助力决策支持智能化、精准化

基于全量数据,借助深度学习、数据挖掘等信息技术,构建“学校、专业、课程、教师、学生”等方面全要素网络化的、具有诊断和预警能力的人才培养质量监控与评价体系,通过领导驾驶舱、综合校情、专业画像、课程画像、师生画像等实时掌握和分析人才培养工作状况和核心指标,通过数据分析判断发展变化趋势,寻找影响因素,推进决策支持从“基于有限个案”向“基于全面数据”转变,推进决策方式从经验型、粗放型向智能化、精准化转变。

3.5 构建数据流通可信机制,保障教育大数据流通安全可靠

教育大数据在流通过程中存在安全性差、防篡改性弱、不可溯源、个人隐私保护机制差、授权管理不可控等问题。区块链作为一项从数字加密领域诞生的新兴技术,其通过将P2P网络、密码学技术、共识机制以及智能合约等多种技术进行深度整合,实现了去中心化、分布式、信息不可篡改的信任机制。分析发现,区块链技术的分布式架构与智能合约技术恰好与大数据环境下数据流通可信需求相吻合,可借鉴区块链在数字加密货币应用及其他领域中的成功经验,构建基于区块链的教育大数据流通可信机制,保障教育大数据流通安全^[5]。

4 结语

大数据为智慧校园的打造及教育信息化建设带来的新机遇,是教育信息化由数字校园迈向智慧校园的基础,也是未来高校竞争发展的核心资产。借助大数据技术,构建教育大数据流通机制,打破各部门、各系统之间的壁垒,实现多源、异构数据的融通共享,提供基于大数据的决策支持,对于深化教育教学改革、提升高校治理水平具有重要价值。

(下转第62页)

形成一个系统集成方案,通过交流讨论等形式分析方案的可行性和实用性,分析关键核心技术中的创新点或可能产生的创新思路,形成反馈用来优化模型,最终实现一个基于机器视觉和自然语言理解相关技术的智能化安防系统设计案例。

2.3 考核阶段,形式灵活多样、评价客观规范

1)采取更为实用灵活的考核方式,抛开以往开卷考试和闭卷考试的范畴,以课程设计项目计划书的形式,编制统一的计划书格式模板。选题上在一定范围内尽可能给学生灵活性和自主性,比如人工智能导论课程要求学生选题应围绕人工智能在生活、工作中的应用,不能脱离实际,而且需要在关键核心技术上体现人工智能算法模型。

2)制定标准规范的考核评价指标体系。对项目计划书从内容完整性、方案科学性、分析深入度、商业可行性等方面综合评判。其中内容完整性占20%(每少一个栏目扣5分,扣完为止),重点考察学生对项目方案制定的总体认识和认真程度;项目背景和研究现状分析占15%,重点考察学生信息检索能力和总结归纳能力,锻炼其问题分析判断能力;核心关键技术分析和解决方案设计占50%,这是课程考查的核心内容,重点是让学生具有较强的算法模型应用和实际解决问题的能力;市场分析和营销策略设计占15%,侧重于学生知识面宽度提升和组织协调能力培养。

3)从项目计划书中挑选部分较优秀或在某一方面具有较强代表性的进行现场交流,采取项目路演的形式,锻炼学生表达能力。通过分析问题,点评成绩,相互学习借鉴,共同进步提高。

2.4 总结阶段,认真分析评价、注重未来发展

结合课程设计项目计划书编制情况,对课程所涉专业领域的发展趋势、未来挑战、应用推广等内容做一个系统梳理和介绍,帮助学生更好地发掘专业兴趣点和未来努力目标,达到教学和育人的双重功效。

3 小结

为提升教学效果、增强学生创新创业能力,文章提出了基于“项目牵引+案例”的混合式课程教学模式改革路径,经过近2个学期的实践检验,教学效果取得较明显提升,但也发现其中存在的一些问题,诸如项目选取适配度、项目计划书撰写质量把控(学生普遍存在对此类文档撰写能力偏低的问题)、与其他课程在教学模式上的差异导致学生思维习惯和认知方式需要时常转换等问题,有待在今后的研究实践中加以解决。

【参考文献】

- [1]赵志强,贾宝柱,许媛媛,安连彤,郝康.项目牵引的轮机工程专业CDIO课程教学模式构建[J].航海教育研究,2020(4):39-44.
- [2]刘细发,夏家莉.基于建构主义指导下的“项目牵引式”教学模式初探[J].电化教育研究,2007(8):65-67.
- [3]卢爱臣,魏建宇,马文彬.以项目为牵引的计算机CDIO创客教学模式研究[J].软件导刊(教育技术),2019(3):31-32.
- [4]张君,赵炼,韦纪英,张雪晶,杨岚,陈芳,周祥文.案例式教学在口腔修复学临床实践教学中的应用[J].中国继续医学教育,2021(25):24-27.
- [5]刘金江,李贺,马晓普,张新刚.面向新工科的人工智能导论实践教学模式初探[J].计算机时代,2021(5):91-93.
- [6]邢胜,左树萍,王晓兰.人工智能导论线上线下混合式教学改革探索[J].黑龙江科学,2021(9):124-125.

(上接第54页)

【参考文献】

- [1]教育部.教育部关于印发《教育信息化2.0行动计划》的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html, 2018-04-18.
- [2]徐蕾,聂峰英.教育大数据在教育管理应用中的研究现状与技术路径[J].江苏高教,2020(12):69-73.

- [3]冯多,李大棚.大数据驱动高校思想政治教育创新的活力、困境及进路[J].现代教育管理,2022(7):113-121.
- [4]张磊.大数据在高校信息化建设应用中的问题与策略[J].科技创新与应用,2017(4):90.
- [5]张人杰,唐雅琳.大数据背景下高校移动学习模式构建研究[J].湖南邮电职业技术学院学报,2020(2):84-86.