

# 数据挖掘在教育大数据中的应用研究

谢磊 龙腾 梁其钰

海南经贸职业技术学院 海南海口 571127

**摘要:**随着大数据时代的到来,海量教育数据的采集、存储、分析、应用成为教育信息化发展的必然趋势。数据挖掘技术在教育领域的广泛应用,为教育决策提供了有力支撑,开启了教育管理与教学模式变革的新路径,当前教育大数据挖掘仍面临诸多现实问题,如数据质量参差不齐、数据安全隐患突出、跨部门数据整合困难、专业人才匮乏等。本文提出加强顶层设计,健全数据管理制度,强化数据安全保护,加快复合型人才培养,推动数据挖掘与教育事业深度融合,更好地服务教育高质量发展。

**关键词:**数据挖掘;教育大数据;应用研究

## 引言

近年来,以云计算、大数据、人工智能为代表的现代信息技术飞速发展,教育信息化进程不断加快,教育领域的结构化、非结构化数据呈爆发式增长,当前我国教育数据总量超过1EB,数据总量和增长速度均居全球前列。面对如此海量的教育数据,传统的分析方法难以胜任,迫切需要引入数据挖掘技术,深入分析数据内在价值,洞察教育发展规律,更好地指导教学实践,服务教育决策。数据挖掘是从大量数据中发现隐含的、先前未知的、潜在有价值的知识的过程,通过机器学习、统计分析、可视化等技术,可从教育大数据中提炼知识模式,预测未来趋势,进而优化教学管理,创新教学模式。

## 一、数据挖掘在教育领域的应用现状

大数据为教育变革带来重大机遇,《教育信息化十年发展规划(2011—2020年)》提出,到2020年基本建成教育决策科学化、管理精细化、教学个性化的人人皆学、处处能学、时时可学的数字化教育新生态。《教育信息化

2023年海南经贸职业技术学院院级科研项目:自由贸易港专项类(重点项目)hnjmk2023402,(基于大数据技术的自由贸易港旅游舆情分析研究)

### 作者简介:

谢磊,男,1984年1月生,海南澄迈县,硕士,副教授,主要研究方向为人工智能技术应用。

龙腾,男,1985年6月生,海南陵水县,硕士,讲师,主要研究方向为计算机科学技术、人工智能。

梁其钰,男,1984年3月生,海南琼海人,本科毕业,副教授,主要研究方向为电子商务、工商管理。

2.0行动计划》进一步强调,要创新教育服务供给方式,深入挖掘数据价值,服务教育教学创新,支撑教育治理能力提升,在政策引领下数据挖掘技术在教育领域的应用不断深入,呈现出蓬勃发展态势。

用于精准画像,为个性化教学提供支撑,通过分析学生的学习行为、学习过程数据,刻画学生个体的知识体系、能力结构,识别学生认知风格、思维特点等,为因材施教、个性化推送学习资源提供数据支持,典型案例如清华大学开发的自适应学习系统,可根据学生的认知特点推送个性化学习内容。优化教学管理,促进精细化管理,利用机器学习算法对教学反馈数据进行实时分析,多维度评估教学质量,基于学生出勤、成绩、就业等数据,对教学管理的薄弱环节进行诊断,为教学过程优化改进提供决策支持,如华中科技大学基于大数据构建的就业质量评价模型,可对就业管理中的不足进行量化评估。创新教育评价,提高评价的科学性,利用聚类、关联、预测等数据挖掘方法,对学生成长的过程性数据进行多元分析,纳入学生的学业水平、创新能力、社会实践等表现,形成既重结果又重过程的综合素质评价体系,如北京十一学校构建的“过程性+结果性”相结合的综合素质评价系统,引入学生活动参与度、获奖情况等非学业数据,拓宽了评价维度。助力教育精准治理,提升治理效能,整合分析校园大数据、政府教育管理数据、互联网教育数据,洞察教育运行态势,及时发现区域教育均衡度、教育质量短板等深层次问题,为政府宏观管理、科学施策提供数据支撑,如武汉市建成的区域教育治理大数据平台,基于全市各校的招生、师资、设施、质量等数据,科学编制教育资源配置规划,推动区域教育均衡发展。

## 二、教育大数据挖掘存在的问题

### (一) 数据质量有待提高

数据质量是开展数据挖掘的前提,教育领域数据来源广泛,采集手段多样,各部门的数据标准不一,数据格式、粒度差异大,导致数据碎片化、非标准化问题突出,不利于后续的分析应用。不同来源的教育数据缺乏必要的校验机制,往往存在数据重复、数据缺失、数据矛盾等质量问题,直接影响数据挖掘结果的准确性。以高考招生数据为例,由于各地上报数据口径不一,统计时间节点不同,一些关键数据项的统计口径存在差异,全国层面数据汇总存在困难,影响对区域教育公平的客观评估,部分数据源缺乏数据治理意识,对“脏数据”的识别清洗不到位,存在较多异常值和离群点,数据可信度不高,数据质量参差不齐已成为制约教育大数据深度应用的瓶颈之一。

### (二) 数据安全与隐私保护不足

教育数据中涉及大量师生的个人隐私信息,在数据采集、存储、流通等各环节极易发生信息泄露,面临较大的数据安全风险。教育部门自身的数据安全意识淡薄,对敏感数据的脱敏处理不到位,管理流程不规范,为不法分子窃取信息提供可乘之机,校外机构违规采集、超范围使用师生个人信息的现象时有发生,一些教育app过度索取手机号码、家庭住址等隐私信息,甚至利用人脸识别等生物信息进行营销,严重侵犯师生的个人隐私。大量含有学生信息的数据集在网上公开售卖,事后监管措施不完善,中消协发布的App个人信息收集与隐私政策测评报告显示,88%的教育类App存在过度收集用户信息的行为。数据挖掘环节缺乏必要的授权和隐私保护机制,风险不容忽视,个人隐私数据的非法收集、非法买卖、非法使用,严重损害师生的合法权益,引发强烈的社会担忧。

### (三) 跨部门数据整合难度大

由于数据分散在不同的业务系统,缺乏统一的数据标准,各部门条块分割、自成体系,数据共享壁垒严重,如高校的就业数据散落在各院系,就业部门难以实时掌握毕业生去向;学生的学籍档案无法与公安户籍数据、人社就业数据联通,影响教育管理的科学性和有效性,这种数据孤岛现象导致宝贵的数据资源沉淀闲置、无法发挥叠加效应。究其原因,缺乏跨部门数据整合的顶层设计,数据标准、数据接口不统一,共享的制度规范缺失;各部门的数据敏感性不同,安全保密要求各异,共享风险难以评估与把控;数据共享的激励约束机制不健全,共享意愿不强,存在本位主义倾向。

### (四) 专业数据挖掘人才匮乏

数据挖掘是一项复杂的系统工程,涉及数学、统计学、计算机科学等多学科知识,对从业人员的专业素养提出很高要求,精通数据挖掘的相关理论、算法和工具,熟练运用机器学习、深度学习等技术对数据进行建模分析。深入理解教育领域的业务特点,洞察教育管理和教学实践的内在规律,才能从海量数据中提炼有价值的关联规则和预测模型。当前,高校培养的数据科学与大数据技术专业人才,缺乏教育学背景,难以胜任教育场景下的数据挖掘工作,而计算机、统计等专业培养的学生,数据分析能力强,但对教育教学规律认识不足,无法准确把握教育数据的特殊性,满足不了教育大数据分析的实际需求,师范院校虽有教育学、心理学等学科优势,但偏重理论研究,实践性、技术性的数据挖掘课程比较欠缺<sup>[1]</sup>。

## 三、深化数据挖掘在教育大数据中应用的对策

### (一) 强化顶层设计,提升教育数据管理水平

数据是数字经济时代的关键生产要素,教育大数据已上升为国家战略资源,党的二十大报告提出,加快建设数字中国,促进数字技术和实体经济深度融合,赋能传统产业转型升级,教育部门要把握数字化发展的时代大势,将教育数据上升到战略高度谋划,加强分类分级管理,建立健全数据标准规范,着力破解数据割裂、碎片化等问题,为数据挖掘应用奠定坚实基础。制定教育数据管理的顶层架构,统筹规划教育数据采集、传输、存储、交换、应用等各环节,明确管理重点,细化管理要求,形成全流程、体系化的数据治理框架,完善教育数据标准体系,优先制定学生学籍、教师信息、招生考试、质量监测等基础数据标准,推动数据规范采集、标准存储,探索建立教育主题数据库、专题数据库、综合数据库,满足不同层次、不同应用场景的数据需求。健全数据质量管控机制,制定教育数据质量管理办法,建立全流程质量监测体系,对采集汇聚的数据进行科学分类,规范管理流程,及时识别数据质量问题,开展数据清洗、数据去重等治理工作,加强对数据提供方、管理方、使用方的质量监管,严格按照数据标准开展数据汇聚、审核、共享、销毁等,确保数据“账相符”“账实相符”,充分运用区块链、隐私计算等技术手段,保证数据来源真实性、传输安全性、访问可控性。创新数据治理工具方法,加快构建智能化的元数据管理平台,实现对数据资源的自动采集、比对、关联、校验,开发可视化的质量评估工具,支持数据缺失、数据异常等问题的实时监测和快速定位,引入数据质量评分机制,建立“红

黄绿”数据质量评级体系，推动数据质量的量化管理和持续改进<sup>[2]</sup>。

## （二）健全数据安全制度，强化隐私保护

大数据安全已上升为国家安全的重要内容，对教育领域而言，师生数据安全事关广大师生的切身利益，事关教育系统的形象声誉，必须高度重视，完善保护，要以网络安全法、数据安全法、个人信息保护法等为依据，加快构建教育数据安全治理体系，强化全流程、多层级的风险防控。细化教育数据分级分类管理制度，制定教育数据分类分级指引，依据数据的重要程度、敏感程度，明确不同类型数据的采集、传输、存储、访问、共享等管理要求。将学生家庭、健康、思想等敏感数据列为重要数据，单独管理，严格授权，防止越权访问，制定面向师生的个人信息收集使用规范，严格限定信息采集的范围、用途和保存时限，最小化使用个人信息，强化对数据处理行为的约束管理，严肃查处超范围收集、非法买卖师生信息等违法违规行为。加强关键环节的技术防护，大力推进数据脱敏技术应用，对师生姓名、身份证号等隐私数据进行加密、置换，提高个人信息保护水平，健全数据库访问控制机制，采用口令、指纹、人脸等多因子认证技术。从物理层、网络层、应用层构建纵深防御体系，严防内外部人员非法访问、篡改数据，加强数据使用全流程监管，利用区块链、同态加密等新兴技术，实现数据流向可追溯、违规操作可问责，大力开展数据安全隐患排查，组织开展数据安全风险评估，及时堵塞管理漏洞<sup>[3]</sup>。

## （三）加强数据融合共享，提高数据整合效率

要充分发挥大数据的聚合效应，必须打通部门数据壁垒，建设资源共享、开放有序、协同联动的大数据应用生态。完善顶层设计，制定教育大数据共享管理办法，创新共享共治机制，统筹推进数据资源的采集汇聚、交换共享、开发利用，建立健全跨部门数据共享的标准规范，研究制定统一的数据采集规程、数据交换协议、数据管理流程，为数据有序共享提供制度保障。搭建大数据共享平台，立足区域实际，整合教育、公安、民政、人社、卫健等部门数据资源，建设覆盖全学段、贯通各领域的教育大数据共享交换平台，运用云计算、虚拟化等技术，对分散的异构数据进行清洗转换、关联整合，实现跨层级、跨部门、跨系统的数据共享与业务协同。建立激励约束机制，完善数据共享绩效评价体系，将数据共享情况纳入各部门年度考核，并与经费投入、人员晋升等挂钩，建立健全数据资产管理制度，明晰数据权属关系，落实数据产权激励，调动各方数据分享的积极性<sup>[4]</sup>。

## （四）加快复合型人才培养，提升数据应用水平

加强复合型数据挖掘人才培养，关系到教育领域数据驱动创新发展的未来，推进产教融合协同育人，支持高校与科技企业、行业部门合作，建设一批大数据联合实验室，开展教育大数据分析应用研究，推动行业专家进课堂、进教材、进系统，加强对学生的实践指导，支持学生到大数据企业实习实训，参与真实项目开发，提升数据挖掘实战能力。完善人才培养课程体系，推动高校面向数字经济发展需求，优化学科专业布局，深化“新工科”改革，加强数据科学专业建设，完善课程设置，强化学科交叉融合，注重理论联系实际，培养学生的数据思维和实践能力，鼓励在计算机、统计等专业开设教育大数据方向，强化教育学背景知识学习，提高学生教育业务的理解，在师范专业培养方案中融入Python、机器学习等数据分析课程，加强技术与教育教学的融合。创新人才评价和激励机制，制定教育大数据人才能力标准，将数据挖掘项目、数据分析成果等纳入高校教师职称评聘、岗位聘用的考察指标，探索实施项目工资、协议工资等分配政策，完善科研助理、技术骨干等配套岗位设置，优化薪酬待遇，稳定数据挖掘团队。

## 结语

数据变革正重塑教育的未来图景，面对日益膨胀的教育数据，如何盘活数据资源宝库，推动数据红利惠及亿万师生，是摆在教育工作者面前的重大课题。必须加强战略谋划和前瞻布局，完善制度规范，健全技术支撑，打通数据壁垒，强化隐私保护，加快人才培养，推动数据挖掘与教育教学、管理决策、科学研究的深度融合，提升教育治理的科学化、精细化、智能化水平，为培养德智体美劳全面发展的社会主义建设者和接班人提供坚实支撑。

## 参考文献

- [1] 王标, 彭瑜, 王维, 等. 数据挖掘在空气质量数值模拟中的应用研究进展[J]. 环境科学研究, 2024, 37(08): 1703-1713.
- [2] 刘涛雄, 尹德才. 大数据在农业经济问题研究中的应用展望[J]. 农业经济问题, 2024, (08): 4-12.
- [3] 王晓妮, 韩建刚. 数据挖掘在教育大数据中的应用研究[J]. 无线互联科技, 2018, 15(18): 135-137.
- [4] 王甲云. 教育大数据在个性化学习路径设计中的应用与挑战——评《教育大数据: 开启教育信息化2.0时代》[J]. 教育理论与实践, 2024, 44(20): F0002-F0002.