

大数据下大学生学习行为预警与干预模型构建

冯伟杰

(汉口学院)

DOI: 10.65285/JYLLYJ.V2I3.004

摘 要: 现在高等教育越来越普及,大学生学业上的问题也变得更加多样、更隐蔽,原来靠期末成绩判断学生是否会挂科的老办法已经不够用了。好在大数据技术发展得快,能帮我们深入分析学生的学习行为,提前发现学业风险。这篇文章就把国内外相关的研究成果整理了一下,从数据来源、评价指标、算法选择、干预办法这四个关键方面,聊聊现有研究的思路和实际应用情况,指出目前存在的不足,再说说未来可能的发展方向,希望能给高校用大数据管理学生学业提供一些参考。

关键词: 大数据; 学习行为; 学业预警; 干预模型; 研究综述

Construction of Early Warning and Intervention Model for College Students' Learning Behavior Based on Big Data

Weijie Feng

Hankou University

Abstract: Higher education is becoming increasingly widespread, and academic problems among college students have also become more diverse and hidden. The old method of judging whether students will fail based on final exam grades is no longer sufficient. Fortunately, the rapid development of big data technology can help us analyze students' learning behaviors in depth and identify academic risks in advance. This paper organizes relevant research findings from domestic and foreign studies, discussing the research ideas and practical applications from four key aspects: data sources, evaluation indicators, algorithm selection, and intervention methods. It also points out existing shortcomings and potential future development directions, aiming to provide some reference for universities in using big data to manage students' academics.

Keywords: Big Data; Learning Behavior; Academic Early Warning; Intervention Model; Research Review

前言

(一) 研究背景

现在高等教育重点在提升质量,可学生学业困难的问题一直影响着人才培养质量。以前判断学生会不会学业出问题,大多看期末考得怎么样,等发现问题时往往已经晚了,而且没法针对性解决。现在教育越来越数字化,高校的教务系统、在线学习平台、校园一卡通这些东西,能记录下学生大量的学习和校园生活数据。这些数据里藏着学生学习状态的规律,正好能用来搭建更科学的预警和干预模型。所以,怎么用大数据找到学习行为和学习成绩的关系,提前发现风险并及时帮忙,成了高校管理中大家都关心的话题。

(二) 研究意义

把国内外的相关研究梳理清楚,明确大数据时代下学业预警和干预模型的关键部分、研究脉络和理论基础,整合不同研究的方法,填补这方面综述类研究的空白,给后面的研究者提供参考。总结现有模型在实际使用中的经验和不足,帮助高校改进学业管理方式,让预警和干预更精准、更有效,让高等教育从“出了问题再补救”变成“提前预防”“个性化指导”。

(三) 国内外研究现状

国外研究学业预警起步早,主要盯着多方面数据的整合和算法模型的优化。有些高校开发了专门的预警系统,收集学生在线学习的参与情况、作业完成情况等过程中的数据,用机器学习算法预测学业风险并进行干预,形成了比较成熟的技术和应用模式。而且国外研究喜欢结合教育心理学、统计学等多个学科的知识,让模型更科学、更实用。

国内这几年相关研究发展很快,主要围绕扩大数据来源、搭建评价指标、优化算法和设计干预办法展开。不过很多研究只针对某一种类型的高校或某个专业,探索适合本土教育情况的预警模型。总体来看,国内研究还存在数据整合不够、评价指标没有统一标准、模型在不同高校适用性不强、干预办法大同小异等问题。和国外比,在结合多学科理论、建立长期有效机制方面还有提升空间。目前大家都认同:把多方面数据深度整合、搭建科学的评价指标、选精准的算法、设计个性化的干预办法,是做好学业预警和干预模型的关键。

一、大学生学习行为预警模型的核心构成

(一) 数据来源与预处理

数据来源类型: 现在研究用的数据早就不只是传统的

考试成绩了,还包括很多过程中的数据和相关数据,主要有四类:一是教务管理数据,比如课程成绩、考勤记录、修了多少学分这些基础学业信息;二是在线学习平台数据,像登录次数、学习多久、作业有没有按时交、测验情况、有没有在论坛互动等;三是校园行为数据,如图书馆借了多少书、在实验室待多久、校园消费规律等,能反映学生的校园生活和学习投入情况;四是主观反馈数据,通过问卷调查、量表测评等方式,了解学生的学习动力、学业压力、学习态度等心理和认知方面的情况。把这些不同来源的数据整合起来,才能更全面地了解学生的学习状态,发现潜在的学业风险。

数据预处理方法:要让数据能用、模型效果好,处理数据是关键步骤。常用的处理方法有:数据清洗,去掉没用的数据、补上缺失的数据、修正异常的数据,让数据更准确;数据标准化,把不同单位、不同范围的数据改成统一标准,避免数据差异影响模型结果;数据脱敏,给学生的隐私信息加密,保证数据使用符合伦理和安全要求;特征选择,通过统计分析、机器学习等方法,挑出和学业风险关系密切的核心数据,减少模型计算量,提高预测效率。

(二) 预警指标体系构建

预警指标就像评价学生学习状态的“尺子”,现在研究大多遵循全面、好操作、有针对性的原则,搭建多层级的指标体系。

一级指标维度:大部分研究都会把核心指标分成三类,即学习投入度、学习成效和学习稳定性。学习投入度看学生在学习过程中参与度和努力程度;学习成效看学生阶段性的学习成果和能力提升;学习稳定性看学生学习行为和成绩的波动情况,这三类指标能全面评价学生的学业状态。

二级与三级指标细化:在一级指标下面,还会拆分成更具体的二级和三级指标。二级指标通常包括在线学习参与度、课堂出勤情况、作业完成质量、测验成绩表现、学分完成进度、学习行为持续性、学习资源利用效率、主观学习状态等;三级指标就是对二级指标的具体量化,比如登录平台的平均次数、出勤率、作业按时提交的比例、单元测验的平均分、核心课程的通过率等。确定每个指标的重要程度时,大多用层次分析法、熵权法等科学方法,一般来说,和学习投入度相关的指标会更重要。

(三) 预警模型算法选择与预警等级划分

算法选择类型:现在研究用的算法主要是机器学习领域的,常见的有逻辑回归、决策树、随机森林、梯度提升树等。不同算法各有优势:逻辑回归原理简单、容易理解,适合基础的线性预测;决策树能处理非线性关系,方便看清数据间的关联;随机森林、梯度提升树等集成学习算法,通过组合多个基础模型,能提高预测的准确性和稳定性,现在很多研究都用这类算法。选择哪种算法,主要看数据特点、研究目标和对模型效果的要求,有些研究还会对比多种算法,选择效果最好的。

模型训练与优化:搭建模型时,通常会把数据分成两

部分,一部分用来训练模型、调整参数,另一部分用来验证模型效果。常用的优化方法有网格搜索法、交叉验证等,通过调整算法参数,让模型预测更准、适用性更强。同时,为了避免模型“学偏”(过拟合),有些研究还会用早停机制、正则化等技术,保证模型稳定可靠。

预警等级划分:根据模型预测的风险概率或得分,结合高校学业管理的实际需求,大多会分成三级预警。一级预警针对学习行为有轻微异常的学生,提醒他们调整学习状态;二级预警针对多项行为指标异常、有一定学业风险的学生,需要针对性指导;三级预警针对学业风险高、可能出现严重学业问题的学生,需要重点干预。预警等级的划分标准,高校可以根据自己的情况调整,确保预警科学实用。

二、个性化干预策略的设计与实践

(一) 干预主体协同机制

要做好干预工作,离不开多个主体的配合。现在研究大多会搭建多主体协同的干预体系,核心参与方包括任课教师、辅导员、学业导师和家长,每个人分工不同、互相配合:任课教师负责课程学习指导,针对学生没学好的知识点进行专项辅导、反馈;辅导员统筹协调干预工作,跟踪学生行为改进情况,提供心理疏导和生活上的支持;学业导师负责制定个性化学习计划,指导学生优化学习方法、做好学业规划;家长配合学校做好日常监督和沟通,形成家校一起育人的闭环。为了让配合更顺畅,有些研究还建议建立跨部门的数据共享平台,打通教务系统、学工系统等的数据壁垒,让干预信息能实时同步、高效传递。

(二) 分级干预策略框架

根据预警等级的不同,现在研究设计了阶梯式的干预策略,保证干预有针对性、有效果。

一级预警干预:主要靠学生自主调整,学校给予支持。比如通过系统给学生推送个性化的学习资源,像慕课视频、习题解析等;组织学习小组,让成绩好的学生分享经验、互相帮助;定期给学生发学习行为分析报告,提醒他们关注自己的异常行为,自己调整学习状态。这个阶段的干预比较轻松,主要鼓励学生主动改变。

二级预警干预:既要指导也要监督,强化外部支持和约束。比如学业导师一对一给学生制定学习改进计划,明确阶段性的学习目标和任务;辅导员定期和学生谈心,了解他们学习上的困难和潜在问题,帮忙解决实际障碍;任课教师增加答疑时间,针对学生没学好的知识点开展专项辅导。这个阶段会定期评估学生的表现,及时调整干预措施。

三级预警干预:进行高强度、全方位的精准干预,重点解决风险问题、提升学生能力。比如要求学生必须参加学业辅导中心的专项培训,强化基础知识和学习方法;心理辅导员给学生做心理健康评估,提供个性化的心理疏导,缓解学业压力和焦虑;建立家校联动的约束机制,加强日常监督和沟通;如果学生是因为专业不合适导致学业困难,就帮忙调整专业或选课方案。这个阶段干预时间较长,会

根据学生的表现动态调整策略。

（三）干预效果评估机制

科学评估干预效果，才能不断优化模型。现在研究的评估体系主要包括三个方面：一是学业指标，比如课程通过率、绩点提升情况、学分完成率等学习成果；二是行为指标，看学生学习行为的改进，比如在线学习时长、作业按时提交率、出勤率等；三是主观指标，通过问卷调查、访谈等方式，了解学生学习动力、学习满意度、自信心等心理状态的变化。评估时大多会做对比实验，设置接受干预的实验组和用传统方式的对照组，分析不同干预方式的效果差异，为优化干预策略提供依据。

三、研究不足与未来展望

（一）现有研究不足

数据整合与质量问题：有些研究只用了单一来源的数据，没有充分整合多方面数据，没法全面反映学生的学业状态；不同研究的数据质量不一样，有些过程性数据的采集标准不统一，影响模型的准确性和可靠性；数据隐私保护和相关伦理规范还需要加强。

指标体系与模型泛化能力：预警指标没有统一的构建标准，不同研究选择的指标差异很大，没法对比；模型大多是针对特定类型高校或专业搭建的，在其他高校或专业适用性不强；算法选择和优化缺乏系统的对比分析，有些模型的原理不好理解。

干预策略与长效机制：干预办法大同小异，没有充分考虑学生的个体差异，比如学习风格、性格特点等；多主体协同干预的效率不高，部门之间协调起来比较麻烦；没有建立长期跟踪和反馈机制，干预效果能不能持续没法保证。

（二）未来发展趋势

数据维度的拓展与深度整合：以后会收集更多类型的

数据，比如学生的社交网络数据、学习环境数据等；加强不同类型数据的融合技术研究，提高数据处理的效率和质量；完善数据隐私保护和伦理规范，在保证数据安全的前提下充分利用数据。

指标体系与模型的优化升级：搭建既有统一标准又能体现个性化的预警指标体系，根据不同高校、专业的特点调整指标的重要程度；引入深度学习等更先进的算法，提高模型的预测精度和适用性；让模型的原理更容易理解，让大家更愿意接受和使用。

干预策略的个性化与智能化：结合教育心理学、认知科学等知识，深入分析学生的个体差异，设计更有针对性的干预策略；利用人工智能技术开发智能干预系统，精准推送学习资源，实时跟踪干预过程；完善多主体协同干预的长期机制，提高干预的系统性和持续性。

跨学科融合与实践应用深化：加强教育科学、计算机科学、心理学、管理学等多个学科的交叉合作，丰富模型的理论基础；扩大研究样本范围，在不同类型高校、不同专业开展研究，提高模型的实际适用性；推动预警干预模型和高校现有的管理系统深度融合，让技术更好地服务于学业管理实践。

参考文献：

- [1] 蒙娇. 基于大数据的学习预警模型及帮扶策略研究[D]. 内蒙古师范大学, 2022.
- [2] 成亚玲, 谭爱平, 胡国安. 基于教育大数据的学习预警信息系统构建研究[J]. 湖南邮电职业技术学院学报, 2022, 21(3): 31-35+71.
- [3] 李利杰, 徐济惠, 张颖. 基于大数据的在线学习精准预警与干预研究[J]. 才智, 2021(29): 64-67.
- [4] 舒莹. 大数据下在线学习危机的精准预警与干预研究[D]. 东北师范大学, 2020.