

在线学习行为数据对学业预警准确性的贡献机制及模型优化

王鹏

重庆师范大学, 重庆 401331

[摘要] 在线学习平台的普及使大规模采集学习行为数据成为可能, 为学业预警系统的精准化建设提供了新的数据基础。然而, 现有学业预警模型在特征选取、动态适应与结果解释三个维度上仍存在明显局限, 制约了预警准确性的进一步提升。本文从行为数据类型与预警机制的关联逻辑出发, 系统梳理在线学习行为数据对学业预警准确性的贡献机制, 将其归纳为参与度识别、学习路径分析、深度学习判断与多维协同四个层次, 并在此基础上提出模型优化的四条路径: 多维特征融合的输入层优化、动态更新机制的引入、可解释性框架的嵌入, 以及干预反馈闭环的构建。研究认为, 行为数据的预警价值不在于数据量的简单累积, 而在于对数据结构与学习规律之间内在关联的深度挖掘; 模型优化的核心目标不仅是提升预测精度, 更在于使预警结果切实服务于教学干预的及时性与针对性, 推动学业预警系统从数据驱动向机制驱动的模式转型。

[关键词] 在线学习行为数据; 学业预警; 贡献机制; 模型优化; 学习分析

DOI: 10. 64635/ja. 2026. 1302

中图分类号: G434

文献标识码: A

Contribution Mechanism of Online Learning Behavior Data to the Accuracy of Academic Early Warning and Model Optimization

Wang Peng

Chongqing Normal University, Chongqing 401331, China

Abstract: The popularization of online learning platforms has made it possible to collect large-scale learning behavior data, providing a new data foundation for the precise construction of academic early warning systems. However, existing academic early warning models still have obvious limitations in feature selection, dynamic adaptation, and result interpretation, which restrict the further improvement of early warning accuracy. Starting from the logical relationship between behavior data types and early warning mechanisms, this paper systematically reviews the contribution mechanism of online learning behavior data to the accuracy of academic early warning, and summarizes it into four levels: participation identification, learning path analysis, deep learning judgment, and multidimensional collaboration. On this basis, four paths for model optimization are proposed: input layer optimization based on multidimensional feature fusion, introduction of a dynamic updating mechanism, embedding of an interpretability framework, and construction of an intervention feedback loop. The study argues that the early warning value of behavior data lies not in the simple accumulation of data volume, but in the in-depth exploration of the internal relationship between data structures and learning patterns. The core goal of model optimization is not only to improve prediction accuracy, but also to ensure that early warning results effectively serve the timeliness and pertinence of teaching interventions, thereby promoting the paradigm transformation of academic early warning systems from data-driven to mechanism-driven.

Keywords: online learning behavior data; academic early warning; contribution mechanism; model optimization; learning analytics

引言

在线学习平台的大规模应用从根本上改变了学习行为的记录方式。传统教育环境中, 学生的学习过程主要通过考试成绩与教师观察加以评估, 信息采集频率低、覆盖维度窄, 难以支撑对学习风险的早期识别。在线学习环境中, 学生的登录记录、视频观看进度、作业提交情况与讨论区发言均以结构化或半结构化的

形式留存于系统日志之中, 为学业状态的实时监测提供了前所未有的数据基础。

学业预警系统的建设正是在这一背景下得到快速推进。然而, 现有预警系统在实践中暴露出若干共性问题: 特征选取多依赖登录频次等浅层行为指标, 忽视行为质量与序列结构所包含的深层信息; 模型以静态训练为主, 难以适应学习行为随课程进度推移的动

态变化；预警结果以风险标签为主要输出形式，缺乏对预警依据的结构化说明，教师难以据此制定有针对性的干预方案。上述问题的根源在于，现有研究对行为数据如何贡献于预警准确性的内在机制缺乏系统性阐释。本文从行为数据类型与预警逻辑的关联出发，系统分析在线学习行为数据对学业预警准确性的贡献机制，并提出模型优化的具体路径。

1 在线学习行为数据与学业预警的理论基础

1.1 在线学习行为数据的类型与特征

在线学习行为数据依据生成方式与信息内涵可区分为三种基本类型。第一类是参与行为数据，包括登录频次、在线时长、视频观看完成率、资源访问次数等，反映学生对学习活动的整体投入程度。这类数据采集成本低、覆盖面广，是当前预警系统中使用最为普遍的数据类型，但仅能呈现学习行为的量的维度，无法直接反映学习质量。第二类是交互行为数据，包括论坛发帖与回帖记录、师生消息往来、协作任务中的贡献记录等，反映学生在社会性学习维度上的参与状态，对预测具有社交孤立风险的学业困难群体具有独特识别价值。第三类是认知行为数据，包括测验答题记录、作业提交时间分布、错误类型分析、重复学习行为等，直接关联学生的认知加工过程与知识掌握状态，信息密度最高，预测效力最为显著，但数据处理难度较大。三类数据在信息维度上具有互补性，多类型数据的系统整合是提升预警准确性的基本前提。

1.2 学业预警的内涵界定与研究现状

学业预警是指通过系统性数据分析，对存在学业失败风险的学生进行提前识别并触发干预机制的教育管理实践，其核心要素包括风险识别的准确性、预警时机的及时性与干预响应的针对性。国内外学业预警研究经历了从经验判断到数据驱动的演进过程，机器学习方法的引入显著提升了预测精度，决策树、随机森林、支持向量机、神经网络等算法均有广泛应用。然而，现有研究仍存在结构性问题：多数研究以预测准确率为核心评价指标，对可解释性与干预可操作性关注不足；模型训练多以学期末数据为基础，早期预警能力有限；跨平台、跨课程的泛化能力普遍薄弱。这些问题共同指向一个核心缺口：行为数据贡献预警准确性的内在机制缺乏系统性阐释。

1.3 行为数据介入预警的理论逻辑

行为数据贡献于学业预警准确性的理论依据来自以下三个核心命题。其一，行为是认知状态的外显标

志，持续的低参与行为、异常的学习时间分布以及反复出错的作答模式，均是认知困难状态的行为映射，具有一定的预测效力。其二，行为模式具有时序结构，学业困难的发生通常经历渐进恶化的过程，在行为层面表现为参与度的持续下滑、交互频次的逐步减少与错误率的阶段性上升，捕捉这一时序结构是实现早期预警的关键。其三，行为数据的预警价值具有维度依赖性，参与行为数据对整体投入风险的识别效力较强，认知行为数据对知识掌握风险的识别效力更为直接，多维数据的协同使用能够覆盖更广泛的风险类型。

2 在线学习行为数据对学业预警准确性的贡献机制

2.1 行为频次数据的预警贡献：参与度识别

行为频次数据主要通过三条路径影响预警准确性。阈值识别通过设定参与行为的最低频次阈值将低于阈值的学生标记为风险群体，方法简单直观，但阈值的合理设定需充分考虑课程类型与学生群体特征，缺乏个体化校准时容易产生误判。趋势分析追踪行为频次随时间推移的变化趋势，将持续下滑的参与行为识别为预警信号，相比阈值识别更能捕捉学业风险的动态发展过程。对比分析将个体行为频次与同课程群体均值进行比较，识别显著偏离群体规律的个体，能够在一定程度上控制课程难度差异对参与行为的影响。行为频次数据的局限性同样不可忽视：频次高不等于质量高，学生可能以频繁但浅层的方式与平台交互，预警系统不能仅依赖频次数据，需引入行为质量维度加以补充。

2.2 行为序列数据的预警贡献：学习路径分析

行为序列数据记录了学生在学习平台上的行为发生顺序与时间间隔，能够反映学习策略倾向与认知加工方式。学业表现良好的学生倾向于遵循系统性的学习路径，表现为资源访问顺序与课程设计逻辑的高度一致与知识点学习的完整覆盖；学业困难学生的学习路径则往往呈现碎片化、跳跃性与回避性三种典型特征。序列模式挖掘、马尔可夫链建模与过程挖掘等分析方法能够从大量行为日志中提取具有统计意义的路径模式，并将个体路径与典型成功路径进行比对，识别路径偏离程度较高的风险学生。行为序列数据的引入，使预警系统从“学了多少”的量的判断扩展至“怎么学”的质的分析。

2.3 行为质量数据的预警贡献：深度学习判断

行为质量数据是三类数据中信息密度最高、预警效力最强的类型，其贡献机制体现在两个层面。知识

掌握状态的直接测量层面,测验答题记录不仅包含对错信息,还涵盖答题时间、修改次数与错误类型等细粒度数据;系统性错误指向概念理解的根本性偏差,随机性错误可能反映注意力分散,二者对应不同的干预策略,在预警层面需加以区分。学习投入深度的间接判断层面,重复学习行为同时包含风险信号(掌握不足)与保护因子(主动应对),预警系统需结合后续测验表现加以综合判断,而非简单归类为负面指标。

2.4 多维行为数据的协同贡献机制

多维行为数据的协同使用是突破单类型数据预警上限的关键路径。协同贡献机制的实现需要解决三个技术问题:特征融合问题,不同类型数据在量纲、分布与时间粒度上存在差异,需通过标准化、降维或特征学习等方法实现有效融合;贡献权重问题,需通过特征重要性分析或注意力机制对各类特征赋予合理权重,避免高频低效特征干扰模型;时序对齐问题,参与行为数据可每日采集,认知行为数据依赖测验与作业的周期性触发,时序对齐是多维数据协同分析的前提条件。

3 现有学业预警模型的局限性分析

3.1 特征选取维度单一

现有预警模型普遍过度依赖参与行为数据,对认知行为数据与交互行为数据的系统性忽视导致两类误判问题:误报率偏高,即将参与度较低但实际学业风险不高的高效自主学习者错误标记为预警对象;漏报率偏高,即遗漏表面参与度较高但深层认知加工不足的真实风险学生。这一问题的形成有其现实原因——认知行为数据的提取需对答题记录进行深度解析,交互行为数据的处理涉及自然语言分析,数据处理复杂度显著高于参与行为数据,制约了研究者对这两类数据的系统性开发。

3.2 模型动态适应性不足

现有预警模型以静态训练为主,在时间、个体与跨课程三个层面均存在适应性不足的问题。时间层面,课程初期与中后期的学生行为模式存在显著差异,以特定时间段数据训练的模型在其他时间段的预警精度往往明显下滑。个体层面,以群体均值为基准的静态模型在处理行为模式差异较大的个体时存在系统性偏差。跨课程层面,不同课程的学习行为规律存在实质性差异,模型直接迁移时性能下降幅度往往超出预期,泛化能力偏弱是现有模型的普遍痛点。

3.3 预警结果可解释性偏低

随机森林、梯度提升树与神经网络等高性能算法

在预测精度上具有明显优势,但其内部决策逻辑对普通用户缺乏透明度。教师面对仅输出“高风险”标签的预警结果,难以判断风险来源,从而无法制定有针对性的干预方案。现有系统虽部分提供特征重要性排序,但这类信息属于模型层面的全局解释,无法反映针对特定学生的个体化风险归因,使预警结果停留于识别层面,难以有效支撑干预决策。

4 基于行为数据的学业预警模型优化路径

4.1 多维特征融合的输入层优化

多维特征融合应遵循三个原则。机制先行原则要求特征选取以行为数据贡献机制的分析为依据,优先纳入在参与度识别、路径分析与深度学习判断三个机制层次上具有明确贡献逻辑的特征,而非以数据可获取性为筛选标准。层次整合原则要求将三类行为数据的特征按信息层次组织:以参与行为特征构建基础信息层,以序列行为特征构建路径信息层,以认知行为特征构建深度信息层,三层特征分别提取后再进行跨层融合,避免直接拼接导致的信息混淆。动态权重原则要求根据课程进度对不同特征赋予动态权重:课程初期参与行为特征权重相对较高,随课程推进认知行为数据逐步丰富,其权重应相应提升,使特征融合策略与数据积累过程保持同步。

4.2 动态更新机制的引入

动态更新机制的实现可从三个技术方向展开。滚动窗口训练以固定时间窗口内的最新数据持续更新模型训练集,淘汰时间距离较远的历史数据,窗口长度需在数据充分性与时效性之间取得平衡。个体基线校正以群体预警模型的基础上为每位学生建立个体行为基线,以个体历史行为的偏离程度作为补充预警信号,有效缓解群体模型对行为模式差异较大个体的预警偏差。迁移学习框架利用已有课程数据训练的预警模型作为新课程预警模型的初始参数来源,通过少量新课程数据进行微调,在数据积累不足的课程初期保持较高的预警能力。

4.3 可解释性框架的嵌入

可解释性框架的嵌入分两个层次展开。全局可解释性层次通过特征重要性分析、部分依赖图等方法揭示不同特征类型对模型整体预测结果的贡献权重,帮助教育管理者理解预警系统的运作逻辑,提升系统在实际中的接受度。局部可解释性层次针对每位被标记为风险的学生,输出驱动该预警结论的具体特征组合及其贡献程度,LIME与SHAP值等方法能够直观呈现特定学生的风险归因,为教师提供具体的干预切入点。

面向教师的预警报告还应将技术性指标转化为教育语言,降低教师理解与使用预警结果的认知成本。

4.4 干预反馈闭环的构建

干预反馈闭环包含四个相互衔接的环节。预警触发环节,预警模型识别风险学生后依据风险等级与风险归因生成差异化预警信息,推送至任课教师、学业导师与学生辅导员,应区分风险等级以避免信息过载。干预执行环节,干预方案的设计应与风险归因相对应,参与度风险对应激励机制调整,认知掌握风险对应补救性教学资源提供,社交孤立风险对应协作学习机会创造。干预效果记录环节,持续追踪目标学生的行为数据变化,记录干预前后关键指标的变化情况,形成结构化的干预效果数据集。模型反馈优化环节,将干预效果数据纳入模型的训练与评估体系,以干预有效性作为模型优化的参考依据,使预警系统从单次预测工具转化为持续迭代的教育改进机制。

5 结语

在线学习行为数据为学业预警系统的精准化建设提供了前所未有的信息基础,但数据的丰富性并不自动转化为预警的准确性。行为数据对学业预警准确性的贡献具有明确的机制逻辑:参与行为数据贡献于参与度识别,序列行为数据贡献于学习路径分析,认知行为数据贡献于深层学习判断,多维数据的协同使用

能够实现风险识别的互补覆盖。理解这一贡献机制,是优化预警模型设计的理论前提。

本文提出的四条优化路径——多维特征融合、动态更新机制、可解释性框架嵌入与干预反馈闭环构建——分别针对现有模型在特征选取、动态适应、结果解释与实践转化四个维度上的结构性局限,四者共同构成完整的优化体系。学业预警系统的建设涉及技术与教育的双重维度,预警系统的伦理边界、数据隐私的保护机制以及教师干预能力的同步提升,均是影响预警系统实际效能的重要因素,有待持续关注与深入探讨。

【参考文献】

- [1] 穆肃,徐欢云.在线学习行为分析研究综述——研究框架与研究发现[J].现代远程教育研究,2019,31(04):72-81.
- [2] 张文兰,刘俊生.学习分析技术在学业预警中的应用研究[J].电化教育研究,2020,41(03):46-53.
- [3] 方海光,高辰璐,张思琦,等.学习分析视角下在线学习行为与学业成绩关系研究[J].中国电化教育,2018(08):94-100.
- [4] 顾小清,顾峰.学习分析的技术框架与实践模式[J].远程教育杂志,2016,34(01):3-11.