

基于改进 iForest 算法的学生行为异常检测研究

王 妍¹ 孙克争¹ 朱亚丽¹
WANG Yan SUN Kezheng ZHU Yali

摘 要

为全面、精准地完成学生行为异常检测，文章将 ReMass 算法与传统 iForest 算法相结合，构建学生行为异常检测系统对学生行为进行异常检测。异常行为检测预警模块中的改进 iForest 算法会根据学生行为数据的质量和密度的分布，确定该行为数据是否存在异常以及异常程度，并给予相应的检测评分。结果表明，ReMass-iForest 综合改进算法评分更接近于人工行为异常评分，且行为异常检测用时在数据样本为 10 000 和 80 000 时，检测用时为 0.2 s 和 0.35 s，用时最短。

关键词

ReMass 算法；学生行为异常；检测系统；检测预警模块；iForest 算法

doi: 10.3969/j.issn.1672-9528.2026.05.039

0 引言

新时代背景下，学校作为促进学生拥有知识技能和社会情感的培育基地，其内部的安全监测决定了学生学习以及生活环境的稳定性^[1]。但随着社会的快速发展，校园霸凌、故意伤害人以及自杀事件多有发生，学生会因为自身内在情感因素、家庭环境因素以及社会与学校环境因素，使得自身压力增加导致日常情绪不稳定、与他人沟通出现障碍、焦虑情绪严重等问题，这些心理问题会严重影响自身与他人安全^[2]。因此通过对学生异常行为的准确检测实现学生心理的及时干预，是目前提高校园安全性的关键途径。

1 基于改进 iForest 算法的检测系统

1.1 系统框架

该系统归纳了所有校园用户种类，并将用户分为管理员、学生、心理咨询师三个类型，且所有用户均可自由登录系统。本文的学生行为异常检测系统拥有用户管理、心理调研、心理检测、线上交流以及异常预警等功能，系统架构如图 1 所示。通过图 1 可知，学生行为异常检测系统前台主要为学生提供权限获取、基础资料管理、学生的心理调研以及学生与学生、学生与老师间的线上沟通等服务。系统的后台服务对象分别为管理员和咨询师，管理员通过后台服务可对系统和角色进行管理，而咨询师通过后台服务，可对学生的海量行为数据

进行管理，并对这些行为数据进行异常检测评分，针对行为异常评分很高的学生，给予异常行为预警，促进学校对学生心理及时、正面地干预^[3]。

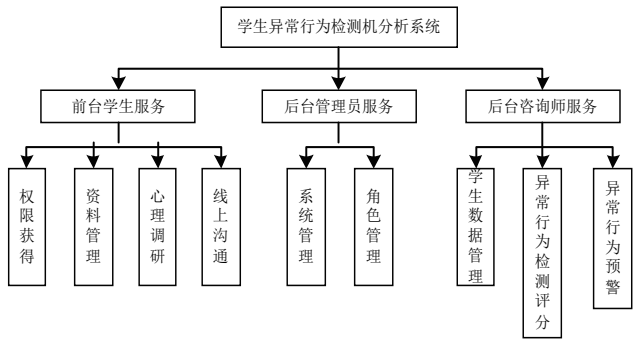


图 1 系统架构

1.2 系统功能模块设计

1.2.1 用户管理模块

该模块用于管理人员管理功能的维护，拒绝所有除管理人员以外的用户的登录，且在管理人员离开系统的一瞬间锁定系统，以防非管理人员入侵操作，且可对管理人员进行添加、删除以及更新等操作^[4]。管理人员在该模块中可完成其个人或管理其他管理员的基本信息管理，包括管理人员的增加、删减、信息的修改、信息的查询等，并可为不同类型的用户开通不同的功能权限，以满足不同类型用户的系统需求。

1.2.2 心理调研测评模块

该模块可支持成人心理压力量表的设置，咨询师可在该模块中设置成人心理压力量表心理测评量表，表中根据需求设置测试题目数量，但过多的题目会引起学生的不耐，因此通常题目数量设置为 20 道，每道题的答案只设置“是”和“否”两个选项，“是”和“否”的分数分别为 1 分和 0 分，可对

1. 江苏商贸职业学院人工智能学院 江苏南通 226011
[基金项目] 江苏商贸职业学院资助课题“基于大数据的学生异常行为检测研究”(SY20231228-17)；南通市青少年工作研究基地立项课题“基于大数据的青少年心理危机预警体系构建研究”(20250048)

学生进行多次测评, 咨询师可根据学生的心理情况进行初步调研^[5]。

1.2.3 异常行为检测预警模块

异常行为检测预警流程如图2所示, 将学生的日常上课行为、购物行为、用餐行为以及社交网络等行为的相关数据录入MySQL数据库中, 异常行为检测预警模块则会从中提取学生的所有行为数据, 通过改进的iForest算法对学生的行为异常情况进行检测评分^[6]。

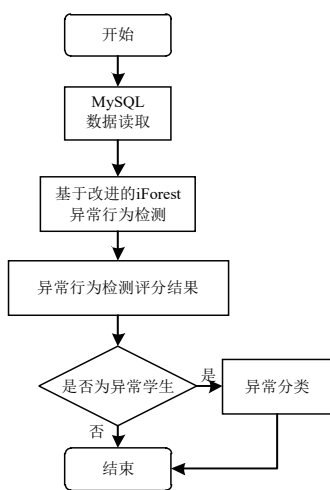


图2 异常行为检测预警流程

2 改进 iForest 算法的学生行为异常检测

2.1 iForest 算法的异常检测

在传统的 iForest 检测算法中, 将超平面沿坐标轴方向分割成多个子空间, 形成多个隔离树 iTree, 此时的 iForest 检测算法因为多个隔离树 iTree 的存在, 形成森林形态。假设隔离树数量为 t 时, 可通过训练 iForest 检测模型后得到的 iForest 森林对异常行为进行判定^[7]。如果某训练数据为 x , 则该数据通过遍历所有隔离树, 最终得到所有测试路径值, 根据路径值打出行异常分值。设定数据节点数量为 n , 则 iForest 森林通过从根节点到所有节点的路径长度的平均值计算行为异常分值的公式为:

$$c(n) = \begin{cases} 2H(n-1) - 2\frac{(n-1)}{n}, & n > 2 \\ 1, & n = 2 \\ 0, & \text{other} \end{cases} \quad (1)$$

式中: $H(n-1)$ 为整个异常分值判定的调和数。训练数据 x 的从根节点到所有节点的平均路径长度计算公式可表示为:

$$E(h(x)) = \frac{1}{e} \sum_{t=1}^e h_t(x) \quad (2)$$

式中: $h(x)$ 表示训练数据 x 遍历第 t 个隔离树期间获取的路径长度; e 表示隔离树数量。为保证异常行为的评分最终的取值范围为 $(0,1)$, 因此需要通过 $c(n)$ 对 $h(x)$ 完成归一化操作, 此时异常行为评分计算公式为:

$$s(x, n) = 2^{\frac{-E(h(x))}{c(n)}} \quad (3)$$

传统的 iForest 检测算法是获取所有节点的路径长度的平均值, 并完成归一化处理后, 实现异常行为评分, 根据分数的大小判定是否需要积极进行心理干预。但该算法在检测学生行为局部异常时, 对于密度相似的正态聚类数据集效

果受限, 导致最终的异常检测结果精准度不高, 因此需要对 iForest 算法进行改进。

2.2 基于 ReMass-iForest 的综合改进算法

为了提高 iForest 检测算法结果的精准度, 本文将 ReMass 算法带到 iForest 算法中, 通过评估学生行为数据的质量和密度的分布, 确定该行为数据是否存在异常以及异常程度, 以此有效分析学生行为的局部异常^[8]。首先通过路径中节点的质量比率得到相对质量, 假设 x 数据局部节点异常分数用 $s_i(x)$ 表示, 路径长度归一化公式为:

$$s_i(x) = \frac{m(\bar{T}_i(x))}{m(T_i(x)) \times \psi} \quad (4)$$

式中: 第 i 个隔离树的叶子节点用 $T_i(x)$ 表示; 最初路径长度和第 i 个隔离树全部节点平均路径长度分别用 $m(\bar{T}_i(x))$ 和 $m(T_i(x))$ 表示; 归一化因子可体现子样本规模用 ψ 表示。

设置一个多元时间序列 $C_{1-20} = (C_1, C_2, \dots, C_{20})$, 该序列长度为 T , 且 C_2 至 C_{20} 均存在特征, 通过概率质量以及质量的比较过程, 对叶子节点的样本进行局部路径长度修改, 此时路径长度归一化公式为:

$$s(x, n_{s,l}) = 2 \times (\ln(n_{s,l} - 1) + \gamma) - \frac{2 \times (n_{s,l} - 1)}{n_{s,l}} \quad (5)$$

式中: n 表示节点数量; γ 表示欧拉-马歇罗尼常数, 该常数不仅可为路径长度的计算过程进行调节, 且可提高对局部行为异常的敏锐度。父节点和子节点之间的相对质量关系公式可表示为:

$$R(m_p, m_l, \psi) = \frac{m_l}{m_p \times \psi} \quad (6)$$

式中: 父节点和子节点质量分别用 m_p 和 m_l 表示, 父节点可表现节点分布特征, 子节点表现节点在整体区域中的重要性。对样本节点的质量进行量化, 可改善对样本节点分布不均区域的异常判断能力。

结合节点质量和路径长度, 调节数据样本所在的节点总路径长度以及父节点的质量比率, 将阈值惩罚机制代入, 其计算过程为:

$$R_2 = d + \frac{\log_2(p)}{\omega \times \log_2(l)} + \log_2\left(\frac{p}{l}\right) \quad (7)$$

式中: 叶节点的样本数量、父节点的样本数量、数据样本所在的节点深度分别用 l 、 p 和 d 表示, 可将数据样本所在的节点深度看成隔离树的高度, 可映射出对节点质量和多维高层质量预测的权重, 同时也是对学生行为异常进行判定的关键环节, 而其中 ω 表示权重参数, 可改变路径长度数据对异常分数的影响程度^[9]。

为了不断更新数据流保证原有特征的稳定性, 可将多尺度窗口带入相邻点差值检测中, 且每条时间序列通过多尺度滑动窗口, 并移动 K 个步长, 此时窗口和行为异常数据周期

之间的叠加性，加大了学生异常行为特征的提取，并在迭代期间，通过异常行为检测算法对学生行为的异常点进行综合评分。

3 实验分析

3.1 数据收集

为了验证本文行为异常检测系统的有效性，本文将从高校学生行为异常检测结果、系统检测效率两个方面进行分析。本文实验所用数据来自某高校的校园一卡通消费记录、图书馆管理系统、网络行为日志、校园教务管理系统等。为了提升整个实验过程的可复现性，需要采用统一的方法对采集到的所有相关数据进行清洗和归一化处理。本文选择基于传统 iForest 算法的行为异常检测系统、基于 K-means 算法的行为异常检测系统、基于 SU-WiForest 算法的行为异常检测系统作为本文基于改进 iForest 算法的检测系统对比项，分析不同系统对学生行为异常检测结果。

3.2 学生行为异常检测结果分析

从校园学生 ID 数据库中选择 8 名学生，并对这 8 名学生的所有行为数据进行提取，邀请班级导师、课任老师、行为咨询师等根据数据情况对 8 名学生进行行为异常检测评分，且设定 0~5 分的检测评分为正常行为，5~7 分学生的检测评分为中等异常行为，7~10 分的检测评分为绝对异常行为。

不同系统评价结果如表 1 所示，可知传统 iForest 算法检测系统在 S408、S710 和 S992 三名学生的异常行为检测中出现了误差，将行为异常的 S408 和 S710 检测为行为中等异常，而将行为中等异常的 S992 检测为行为异常；K-means 算法检测系统将行为中等异常的 S157 和 S992 检测为行为异常，将行为异常的 S408 和 S710 检测为行为中等异常；SU-WiForest 算法检测系统和本文改进 iForest 算法检测系统的检测评分结果均很理想，检测评分结果均与学生行为人工评价结果一致。但相对而言，改进 iForest 算法检测系统的行为异常检测评分更接近于人工行为异常评分，检测结果最为准确。

表 1 不同系统评价结果

学生 ID	人工行为异常评分	评价结果	传统 iForest 算法	基于 K-means 算法	基于 SU-WiForest 算法	本文基于改进 iForest 算法
S105	3.35	行为正常	4.51	4.82	3.17	3.28
S157	6.49	行为中等异常	5.23	7.74	6.20	5.51
S229	2.73	行为正常	4.02	3.36	2.99	2.75
S303	3.59	行为正常	4.78	4.61	3.22	3.56
S408	8.74	行为异常	6.93	6.76	8.36	8.71
S663	3.68	行为正常	3.04	5.72	3.82	3.72
S710	9.21	行为异常	6.95	6.73	9.03	9.17
S992	6.02	行为中等异常	7.46	7.72	6.27	6.04

3.3 系统行为异常检测效率分析

图 3 为学生行为异常检测效率。检测用时最长的是 K-means 算法检测系统，其随着数据样本的增加，检测用时呈现上升趋势，当数据样本达到 80 000 时，检测时间达到最高点 1.6 s。传统 iForest 算法检测系统对于异常点样本的检测从样本数量 20 000~80 000 之间时检测用时缓慢上升，在数据样本为 80 000 时，检测时间最高点为 0.96 s。SU-WiForest 算法检测系统虽然在整个检测过程中用时呈缓慢上升趋势，但在数据样本达到 80 000 时，最高检测用时只有 0.46 s。所有系统中，本文基于改进 iForest 算法的检测系统检测用时最短，数据样本 10 000 时，检测用时为 0.2 s，数据样本为 80 000 时的最高检测用时 0.35 s，均为所有系统中最低。由此可见本文采用改进 iForest 算法下的检测系统对学生行为异常检测，不仅检测结果精准度最好，而且检测效率最高。

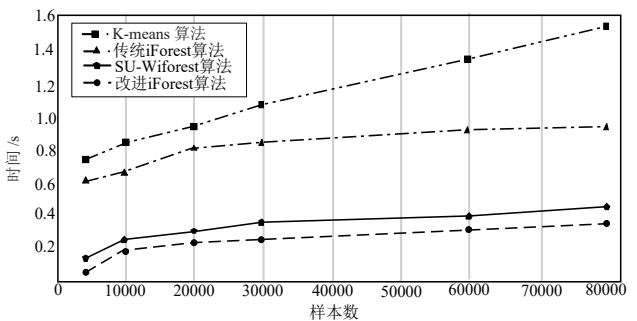


图 3 学生行为异常检测效率

4 结论

本文提出基于改进 iForest 算法的检测系统，对比分析发现传统 iForest 算法检测系统、K-means 算法检测系统均会出现行为异常检测偏差，造成行为异常的学生被检测为行为中等异常，或是行为中等异常的学生被检测为行为异常，严重影响后续学生心理干预方案的设计以及最终的干预效果，而只有本文 ReMass-iForest 的综合改进算法检测系统检测结果与人工检测结果极为接近。在检测效率分析中，所有系统的检测用时均随着数据样本数量的增加呈现上升趋势，但只有基于改进 iForest 算法的检测系统在数据样本 10 000 时，检测用时为 0.2 s，数据样本为 80 000 时的最高检测用时 0.35 s，均为所有系统中最低。因此，本文研究内容实现了高校学生行为异常的精准、高效率检测，为未来该方向的研究打下坚实的理论基础。

(下转第 175 页)

and Education: Artificial Intelligence, 2023, 5: 100177.

[23] 褚乐阳, 刘泽民, 王浩, 等. 大模型支持的教师循证实践: 行动框架与案例应用 [J]. 开放教育研究, 2024, 30(4): 91-103.

[24] 卢宇, 余京蕾, 陈鹏鹤. 基于大模型的教学智能体构建与应用研究 [J]. 中国电化教育, 2024(7): 99-108.

[25] 余胜泉, 熊莎莎. 基于大模型增强的通用人工智能教师架构 [J]. 开放教育研究, 2024, 30(1): 33-43.

[26] 徐磊, 胡亚豪, 潘志松. 针对大语言模型的偏见性研究综述 [J]. 计算机应用研究, 2024, 41(10): 2881-2892.

[27] GLM Team. ChatGLM: a family of large language models from GLM-130B to GLM-4 all tools[PP/OL].(2024-07-30) [2026-05-07].<https://doi.org/10.48550/arXiv.2406.12793>

[28] Wang Weihang, Lü Qingsong, Yu Wenmeng, et al. CogVLM: visual expert for pretrained language models[PP/OL]. (2024-02-04)[2026-04-07].<https://doi.org/10.48550/arXiv.2311.03079>.

[29] 智谱 AI. 智谱 AI 推出新一代基座模型 GLM-4[EB/OL]. [2024-09-07]. <https://zhipuai.cn/devday>.

[30] 刘燕, 荆国茹, 李玉玲. 师范生全学期教育实习评价指标体系构建: 以内蒙古师范大学生物科学专业为例 [J]. 教育教学论坛, 2022(25): 164-168.

[31] Chen Caixia, Chen Jun, Shi Chun. Research on credit evaluation model of online store based on SnowNLP[C]//International Conference on Intelligent Automation, Energy and Environment Engineering (ICAEER 2018). Les Ulis: EDP Sciences, 2018: 03039.

[32] 张小艳, 闫壮. 融合大语言模型的三级联合提示隐式情感分析方法 [J]. 计算机应用研究, 2024, 41(10): 2900-2905.

[33] Zhang Tianyi, Kishore V, Wu F, et al. BERTScore: evaluating text generation with BERT[PP/OL].(2020-02-24)[2026-05-07]. <https://doi.org/10.48550/arXiv.1904.09675>.

[34] Lu Sicheng, Liu Qiang, Zhang Zhikui. Sentiment Analysis of Weibo Platform Based on LDA-SnowNLP model[C]//2023 2nd International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM). Piscataway: IEEE, 2023: 505-510.

【作者简介】

杨怡(2000—), 女, 云南大理人, 硕士研究生, 研究方向: 智慧教育、计算机视觉。

张洪源(2000—), 男, 河北秦皇岛人, 硕士研究生, 研究方向: 智慧教育、计算机视觉。

焦方旭(2000—), 男, 山东济南人, 硕士研究生, 研究方向: 智慧教育、计算机视觉。

赵姗(1995—), 女, 云南昭通人, 博士, 讲师, 研究方向: 计算机视觉、图像处理。

高丽金(1981—), 通信作者(email:gaolijin@ynnu.edu.cn), 女, 云南曲靖人, 硕士, 讲师, 研究方向: 模式识别、图像处理。

杨扬(1980—), 男, 云南昆明人, 博士, 研究方向: 计算机视觉、模式识别。

(收稿日期: 2025-10-13 修回日期: 2026-05-11)

(上接第 165 页)

参考文献:

[1] 徐丽君. 智慧校园学生行为数据采集与异常检测平台 [J]. 长江信息通信, 2025, 38(8): 102-104.

[2] 谭暑秋, 汤国放, 涂媛雅, 等. 教室监控下学生异常行为检测系统 [J]. 计算机工程与应用, 2022, 58(7): 176-184.

[3] 廖鹏, 刘宸铭, 苏航, 等. 基于深度学习的学生课堂异常行为检测与分析系统 [J]. 电子世界, 2018(8): 97-98.

[4] 任孟其, 宋汝鑫, 王萌, 等. 面向智慧校园的学生异常行为检测 [J]. 哈尔滨师范大学自然科学学报, 2017, 33(3): 20-24.

[5] 余建, 林志兴, 肖香梅. 采用分形维数聚类算法的云计算用户异常行为检测方法 [J]. 三明学院学报, 2025, 42(3): 27-38.

[6] 李晨琛, 窦敬, 孙霁云, 等. 基于姿态表征的轻量化行人异常行为检测算法 [J]. 西安邮电大学学报, 2025, 30(3): 94-102.

[7] 蔡佳炜, 李建文, 谢胜达. ReMass-iForest+: BDS 广播星历异常检测方法 [J]. 导航定位学报, 2025, 13(3): 85-98.

[8] 刘一宁, 陈柏安, 杜鹏程, 等. 基于 MDLOF-iForest 和 M-KNN-Slope 的公共建筑负荷异常数据识别与修复 [J]. 综合智慧能源, 2025, 47(3): 62-72.

[9] 李雨笑, 胡居荣, 邢延潇, 等. 基于改进孤立森林的海面小目标检测方法 [J]. 雷达科学与技术, 2024, 22(6): 628-636.

【作者简介】

王妍(1993—), 女, 江苏南通人, 硕士研究生, 讲师, 研究方向: 大数据分析。

孙克争(1983—), 男, 江苏赣榆人, 硕士研究生, 副教授, 研究方向: 算法优化。

朱亚丽(1987—), 女, 江苏南通人, 硕士研究生, 副教授, 研究方向: 人工智能。

(收稿日期: 2025-09-22 修回日期: 2026-05-08)