

基于校园学习行为与主观感知数据的学生画像分析

纪志荣, 路春燕, 胡洁莹

福建农林大学 计算机与信息学院, 福建 福州 350002

DOI:10.61369/ASDS.2025040015

摘要： 学生校园行为数据日益成为刻画学习过程与评估学业成效的重要依据，本文基于“以学生为中心”的教育理念，以某高校在读学生为研究对象，借助社会认知理论与教育数据挖掘方法，从手机使用与网络依赖，心理健康与压力感知，学习投入与自我效能三个核心维度进行问卷设计与调查，采集其客观行为数据与主观感知数据，基于K-mode聚类算法识别学生群体行为特征，构建多维学生画像体系；并引入决策树分类模型，以综合测评成绩排名为标签变量，挖掘行为特征与学业表现之间的路径关联。研究有助于推动高校学生管理由“被动风险识别”向“主动成长引导”转型，对提升高等教育质量具有重要应用价值。

关键词： K-mode 聚类；学习投入；画像分析；决策树；主观感知数据

Student Profiling Analysis Based on Campus Learning Behaviors and Subjective Perception Data

Ji Zhirong, Lu Chunyan, Hu Jieying

College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002

Abstract： With the growing importance of behavioral data in educational contexts, campus behavior has emerged as a critical indicator for understanding learning processes and evaluating academic performance. Anchored in the "student-centered" educational paradigm, this study investigates undergraduate students at a Chinese university by integrating Social Cognitive Theory with educational data mining techniques. A multidimensional survey instrument was developed around three core domains—mobile phone usage and internet dependence, mental health and perceived stress, and learning engagement and self-efficacy—capturing both objective behavioral metrics and subjective self-reported perceptions. Utilizing the K-modes clustering algorithm, the research identifies distinct student behavioral typologies and constructs a comprehensive multidimensional student profiling framework. Additionally, a decision tree classification model is employed, using comprehensive academic evaluation rankings as the outcome variable, to uncover the associative pathways between behavioral features and academic achievement. The result provides empirical support for transitioning student management in higher education from reactive risk identification to proactive developmental guidance, offering significant implications for the enhancement of educational quality and student support systems.

Keywords： K-mode clustering; learning engagement; profile analysis; decision tree; subjective perception data

引言

在推进教育高质量发展的背景下，我国高等教育人才培养正由“规模扩张”转向“内涵提升”，教育质量评价体系逐步从结果导向向过程导向转型。如何精准识别大学生学习状态并提供动态支持，已成为当前高等教育改革的关键议题。

传统学生管理主要依赖成绩、课程完成等静态指标，难以及时发现学生在学习过程中的心理波动和行为偏差，导致干预滞后、效果有限。随着学习分析（Learning Analytics）和教育数据挖掘（Educational Data Mining）技术的发展，数据驱动的“教育画像”（Educational Profiling）成为高校实现学生精细化管理与个性化支持的重要工具，为教育治理从“静态评估”向“动态识别”转型提供

基金项目：

2020年度福建省教育科学“十三五”规划课题“基于教育大数据的学习分析与干预机制的实证研究——以X大学为例”（FJJKCG20-217）；

2022年福建农林大学本科教育教学改革项目：“专业思政”视阈下统计类核心课程思政改革与建设研究”（111422140）。

作者简介：纪志荣（1979-），女，汉族，河北石家庄人，硕士，讲师，研究方向：应用统计学。

支撑。

当前教育画像研究多聚焦于客观行为数据建模，如签到、图书借阅、学习轨迹和消费记录，借助聚类、分类、预测等算法识别学生群体特征和学业风险^[1-3]。如翟鸣宇^[4]等基于学生基本信息与消费数据构建聚类标签，万安庆^[5]通过成绩、一卡通数据等建立学生亚健康画像。但多数研究偏重量化行为，忽视学生学习过程中的主观感知，缺乏对学习动机、自我效能、压力感知和网络依赖等心理维度的整合。

教育心理学研究表明，学习动机、情绪状态、自主感和网络使用体验等主观变量，深刻影响学生的行为选择与学业持续性。相关实证也表明，压力升高、幸福感下降和网络依赖，往往先于成绩滑坡，构成重要预警信号^[8-9]。如吴绮琳等^[10]发现高频手机使用会削弱注意力，引发焦虑，彭静雯^[11]强调感知自主性对学习内驱力的重要性，姚丽^[12]借助网络行为日志预测学业表现，王媛^[13]从目标设定与自我适应角度探讨学习机制。

因此，亟需构建融合主观感知与行为特征的教育画像，全面反映大学生学习生态与发展潜力，推动学生管理从“事后处理”向“过程干预”转变，为高校构建支持性、发展型培养机制提供理论基础与实践路径。

一、研究设计

（一）分析框架

基于“以学生为中心”的教育理念为指导，融合社会认知理论与教育数据挖掘方法，以在读大学生为研究对象，采集其校园行为数据与主观感知问卷数据，聚焦学习投入、网络依赖、心理压力与自我效能等核心变量，采用K-mode聚类算法对学生群体进行分类，构建多维度的学生画像体系。并进一步引入决策树分类模型，学业表现以学生综合测评成绩排名作为标签变量，挖掘学生不同行为一感知特征与学业表现之间的路径关系，构建具备可解释性的行为一成绩关联规则模型。最终提出基于学生行为特征的差异化干预策略，推动高校学生管理工作由“被动风险识别”向“主动成长引导”转型，为高等教育质量提升提供实证基础与政策参考。

（二）研究对象

为了了解大学生校园行为特征，选取A高校理工科统计学专业在读学生为调查对象，通过课堂问卷发放方式，共回收有效问卷508份，有效回收率为95.13%。其中，男生183人（36.02%），女生325人（63.98%）；农村生源348人（68.50%），城市生源160人（31.50%）；独生子女134人（26.38%），非独生子女374人（73.62%）；大一至大四学生分别为89人（17.52%）、194人（38.19%）、152人（29.92%）、73人（14.37%）；担任学生干部的有233人（45.87%）。

（三）研究工具

本研究基于文献综述与理论框架，自编调查问卷，采用李克特五点量表评分法，题项分为有利项与不利项，按“完全不同意”至“完全同意”赋值1-5分。问卷包括个人信息采集及课程体验、学习投入和学习收获三大量表，围绕“手机使用与网络依”“心理健康与压力感知”“学习投入与自我效能”三个核心维度展开，旨在全面刻画学生的心理、行为与认知状态。数据分析采用SPSS 26.0，首先对各子量表和维度进行信效度检验，随后采用主成分分析和最大方差旋转法进行探索性因子分析，以特征值大于1为因子提取标准，剔除因子负荷低于0.5的题项，确保量表

结构的科学性与信效度。

1. 手机使用与网络依赖量表

参考文献^[13-14]本量表围绕开始使用手机的年龄、平均每月手机娱乐消费、平均每天手机使用时长、手机依赖程度、短视频依赖程度以及网络使用情境（如上课、寝室、夜间）六个维度设计，共包含15道题目。信度分析结果显示，量表的Cronbach's α 系数为0.804，表明量表具有良好的内部一致性。因子分析结果表明，KMO值为0.895，Bartlett球形检验的p值小于0.001，总方差解释率为81.23%。结果说明该量表各部分题项设计合理，结构效度良好，整体信效度达到较高水平。

2. 心理健康与压力感知量表

参考文献^[7]，量表从睡眠行为、未来焦虑情绪状态、人际关系与社会支持、学业压力及就业压力五个方面构建，共包含18道题目。信度检验显示，量表Cronbach's α 系数为0.882，KMO值为0.895，Bartlett球形检验p值小于0.001，累计解释总方差变异率为82.19%。结果表明，本量表设计合理，具有较好的结构效度和内部一致性，能够有效测量大学生的心理健康与压力感知水平。

3. 学习投入与自我效能量表

参考文献^[7, 9]，量表依据课堂投入、课外投入、学习计划、学习交流、学习动力及学习反思六个维度设计，共计20道题目。经信度检验，Cronbach's α 系数为0.794，KMO值为0.864，Bartlett球形检验结果p值小于0.001，总方差解释率达到81.23%。结果表明，量表具有良好的内部一致性与结构效度，能够较为准确地评估学生的学习投入水平与自我效能感。

二、学生教育画像分析

（一）k-modes 聚类理论

Huang等人^[15]提出了适用于分类数据的k-modes聚类，该聚类采用SMD（Simple Matching Distance）方法实现对分类变量的处理，用众数代替均值，用Hamming距离计算两个样本点之间的距离，其目标函数定义为样本数据点与其对应聚类中心的距

离总和。

设 $X = \{x_1, x_2, \dots, x_n\}$ 为样本点集合, 样本 $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 的属性为 $\{A_1, A_2, \dots, A_m\}$, 对于属性 A_i 的取值为 $Dom(A_i) = \{a'_i, a''_i, \dots, a''_i\}$ ($l > 2$), 任意两个样本点 x_i 和 x_j 的 k-modes 距离为:

$$D(X_i, X_j) = \sum_{l=1}^m d(X_{il}, X_{jl}), \text{ 其中, } d(X_{il}, X_{jl}) = \begin{cases} 0, & X_{il} = X_{jl} \\ 1, & X_{il} \neq X_{jl} \end{cases}$$

根据轮廓系数选择合适的类数, 聚类簇数的选择关系到聚类结果的优劣, 轮廓系数结合了凝聚度和分离度, 适合用于评估聚类效果的好坏, 轮廓系数定义为:

$$S = \frac{\sum_{i=1}^n \frac{b_i - a_i}{\max\{b_i, a_i\}}}{n}$$

其中, a_i 表示第 i 个样本与其同一簇群中所有其他数据之间的平均距离, 即量化簇内的凝聚度; b_i 表示其与最近一个聚簇的平均距离, 用于量化簇内的分离度; S 表示全部样本轮廓系数的平均值; 当 S 小于 0 时, 表示 S 与其簇内元素的平均距离大于最近的其他簇, 聚类效果不好; 反之, 当 S 越趋近于 1 时, 表示聚类效果越好。

(二) 学生画像分析

对问卷数据进行聚类分析, 当聚类个数 $K=6$ 时, 如图 1 所示, 聚类效果最佳。通过对六类学生群体的特征标签进行标注和分析, 揭示各类群体在性别、年级分布、学习行为及心理状态等方面的典型特征, 并绘制词云图 (1-6)。

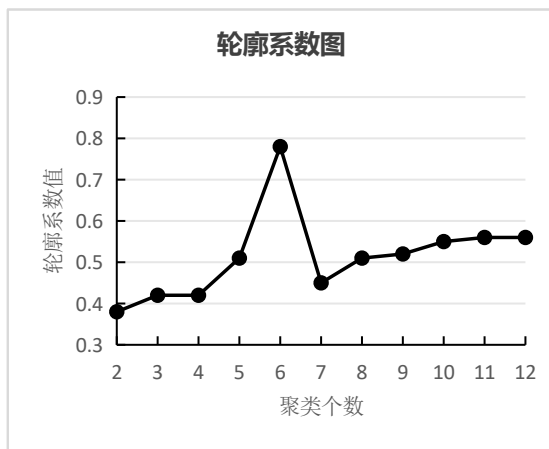


图 1

第一类: “自律高效型学生” (词云图 1), 共 56 人, 男生人数为女生的四倍, 主要集中在大二和大三年级。该群体以农村男生为主, 多担任班干部, 手机依赖度低, 学业及就业压力较小。学习动力强, 能严格执行学习计划, 每周去图书馆 1 至 3 次, 课堂专注时间超过 30 分钟, 晚间 11 点至 12 点入睡, 保证 7 至 8 小时睡眠。该群体几乎无挂科或重修记录, 表现出较高的自律性和学习效率。

第二类: “积极进取型学生” (词云图 2), 共 197 人, 女生占 76%, 涵盖大一至大四学生, 以大二大三为主。大部分来自农村, 非独生子女。作息较晚, 通常在午夜至凌晨 1 点入睡, 睡眠时长 6 至 7 小时。每周图书馆或自习室学习 1 至 3 次, 与师生交流频

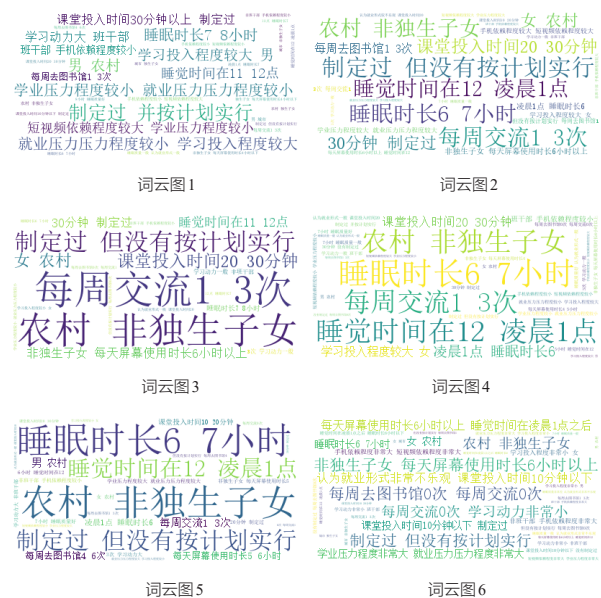
率为 1 至 3 次。课堂集中时间约 20 至 30 分钟, 虽有学习计划但执行不完全。该群体学业表现中上, 学习积极性较强。

第三类: “学习投入不足型学生” (词云图 3), 共 69 人, 男女比例约 1:3, 集中在大二二年级, 上一学年挂科率达 25%。以农村非独生子女为主, 每周与师生交流 1 至 3 次, 课堂投入 20 至 30 分钟。虽有学习计划, 但执行力较弱。睡眠时间在 11 点至 12 点, 日均屏幕使用超过 6 小时。该群体自我管理较弱, 学习投入不足。

第四类: “晚睡克制型学生” (词云图 4), 共 63 人, 性别均衡, 大三学生占比最高, 上一学年挂科率高达 48%。主要为农村非独生子女, 作息较晚 (12 点至凌晨 1 点), 睡眠 6 至 7 小时。每周与师生交流 1 至 3 次, 图书馆或自习室学习频次最高, 达 4 至 6 次, 手机使用较克制, 日均 5 至 6 小时。该群体虽晚睡, 但学习行为较为有序。

第五类: “晚睡低效型学生” (词云图 5), 共 77 人, 男女比例约 1:2, 大二人数最多。挂科或重修三科以上人数较多。该群体晚间 12 点至凌晨 1 点入睡, 睡眠 6 至 7 小时, 课堂投入时间较多 (20 至 30 分钟), 但学习效果不佳, 成绩不理想, 表现为学习效率低下。

第六类 “焦虑放任型学生” (词云图 6), 共 46 人, 女生占多数, 未涵盖大四学生。每学期挂科 1 至 2 门, 几乎不去图书馆, 与师生无交流, 课堂投入不足 10 分钟。尽管设有学习计划, 但未能执行。该群体感受较大学业和就业压力, 对未来悲观, 手机及短视频依赖严重, 缺乏自我调节能力, 学习状态较差。



三、学生校园行为与学业表现的关联性分析

综合测评作为衡量学生综合素质的重要方式, 其结果不仅反映学生在德、智、体等方面的真实表现, 也对学生学业发展产生导向作用。基于决策树算法原理, 通过信息增益 (Information Gain) 与基尼指数 (Gini Index) 评估特征变量的重要性, 并筛选关键行为指标。综合测评排名被划分为四个等级: “高水平 (前 25%)” “中上水平 (25% - 50%)” “中下水平 (50% - 75%)”

和“低水平（后25%）”，作为分类标签训练模型。选取样本量不少于10的主要分类路径，提取并分析对应的行为分类规则。结果显示，除入睡时间、睡眠质量、就业压力程度与每周图书馆使用频率外，其他行为变量与综测成绩存在较强相关性（表1），具体分类规则：

（一）学习投入 $\in$ [3,4]，学习计划 $\in$ [1]，学习动力 $\in$ [3,5] $\rightarrow$ 综测排名小于25%；

（二）学习投入 $\in$ [3,4]，学习计划 $\in$ [2,3]，工作预测 $\in$ [1,3] $\rightarrow$ 综测排名小于25%；

（三）学习投入 $\in$ [1,2]，每天手机使用时间 $\in$ [2,4]，每周交流学习次数 $\in$ [2,4] $\rightarrow$ 综测排名在25%–50%之间；

（四）学习投入 $\in$ [1,2]，每天手机使用时间 $\in$ [2,4]，每周交流学习次数 $\in$ [1]，学业压力程度 $\in$ [2,4] $\rightarrow$ 综测排名在50%–75%之间；

（五）学习投入 $\in$ [1,2]，每天手机使用时间 $\in$ [2,4]，每周交流学习次数 $\in$ [1]，学业压力程度 $\in$ [1] $\rightarrow$ 综测排名大于75%。

表1 校园行为特征与成绩关联系数

特征名	系数	特征名	系数
手机使用时长	0.0814	课堂投入	0.2241
手机依赖程度	0.0415	学习计划	0.0976
短视频依赖程度	0.0403	每周图书馆次数	0.0596
入睡时间	0.0865	每周交流学习次数	0.0905
就业形式预测	0.0611	学习动力	0.1517
学业压力程度	0.0667		

四、结论与建议

本文基于 K-mode 聚类与决策树分类方法，实证分析了大学生校园行为数据与综合测评成绩的关联，识别出六类典型学生群体：自律高效型、积极进取型、学习投入不足型、晚睡克制型、晚睡低效型和焦虑放任型。研究发现，学生在学习投入、交流互动、作息规律及手机使用等行为维度存在显著差异，且与学业成绩紧密相关。

针对不同类型学生，提出差异化干预建议：对自律高效型学生，鼓励其强化自主学习与目标导向，提供科研、竞赛等高阶成长平台，发挥其榜样作用，促进良性学习氛围；对积极进取型学生，强化时间管理与计划执行能力，推动学术交流与阶段性激励，提升持续学习动力；对学习投入不足型学生，注重行为认知与时间管理训练，实施结构化辅导与朋辈支持，促进行为改善；对晚睡克制型学生，帮助调整学习时间分布，强化学习策略反思，配合导师指导，实现努力与成果有效衔接；对晚睡低效型学生，开展学习策略与效率提升辅导，加强睡眠教育，通过榜样激励促使习惯转变；对焦虑放任型学生，则优先开展心理支持与个别干预，辅以行为分解与数字戒断训练，重建学习信心与行动力。

综上，大学生校园行为不仅反映其学习态度与生活习惯，更与学业成绩密切相关。通过典型行为画像的识别与针对性干预，有助高校实现从经验驱动向数据驱动的精准管理与帮扶。未来建议结合行为时序与心理测评数据，构建动态预警机制，推动形成“识别—干预—反馈”闭环式育人体系，从而更好地服务于学生全面发展与高质量成长。

参考文献

[1] 余明华, 张治, 祝智庭. 基于可视化学习分析的研究性学习学生画像构建研究 [J]. 中国电化教育, 2020, (12): 36–43.

[2] 张紫微, 林子涵, 廖会心, 等. 面向学习成功的非认知能力学生画像可视化 [J]. 现代教育技术, 2021, 31(12): 94–102.

[3] 王惠惠, 董永权, 和文斌, 等. 基于聚类集成的学生群体画像构建 [J]. 江苏师范大学学报 (自然科学版), 2022, 40(03): 46–50.

[4] 翟鸣宇, 程建, 王苏桐, 等. 基于 K-prototype 聚类的大学生教育画像分析 [J]. 大连理工大学学报 (社会科学版), 2021, 42(06): 22–31.

[5] 万安庆. 智慧校园环境下大学生亚健康数据挖掘方法应用研究 [D]. 南京: 南京中医药大学, 2017.

[6] 刘智, 康令云, 刘三女牙, 等. 学习者社会网络交互、情绪表征与学习成效的关系研究 [J]. 中国远程教育, 2020, (06): 31–39+59+77.

[7] 刘佳. 大学生学业压力、学业自我效能感和学业拖延的关系及干预研究 [D]. 云南师范大学, 2024.

[8] 张旭东, 马塘生, 李清. 大学生网络成瘾对学业挫折感的影响: 心理生活质量的中介作用 [J]. 心理学探新, 2022, 42(04): 356–361.

[9] 于倩, 刘金兰, 赵远. 大学生学习动机对学习参与及学业成就的影响研究 [J]. 大连理工大学学报 (社会科学版), 2018, 39(06): 100–106.

[10] 吴绮琳, 刘世超, 黄时华, 等. 特质性心流与网络游戏成瘾: 学业自我效能感的中介作用 [J]. 中国健康心理学杂志, 2023, 31(09): 1407–1412.

[11] 彭静雯. 大学生学习内驱力影响因素及其作用机制——基于扎根理论的分析 [J]. 现代大学教育, 2023, 39(02): 93–102.

[12] 姚丽, 崔超然, 马乐乐, 等. 基于校园上网行为感知的学生成绩预测方法 [J]. 计算机研究与发展, 2022, 59(08): 1770–1781.

[13] 王媛, 刘阳. 自我选择与环境塑造: 本科生学习投入的养成逻辑——基于 H 区高校本科生的实证考察 [J]. 当代教育与文化, 2020, 12(05): 106–116.

[14] 荣婷. 手机依赖度对大学生身心健康、人际关系、学习状态的影响研究——基于全国 2240 所高校调查的实证分析 [J]. 黑龙江高教研究, 2018, 36(06): 114–118.

[15] Zhexue H. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values.[J]. Data Min. Knowl. Discov. 1998, 2(3): 283–304.