

数智赋能学生特征识别与教学改进策略研究

钟秋杨 张宛静*

重庆轻工职业学院智能制造学院

摘要: 在大数据、人工智能时代背景下, 学生学习成效和内在发展需求已无法从分数这一单一指标中得到, 现代校园各项学生生活和学习数据需要被认识并利用。本文基于高校学生学习行为与成效数据, 改进 K-means 聚类算法, 构建融合多维数据的聚类模型, 实现学生学习特征精准分类, 并根据分类特征提出有针对性教学策略, 为教育评价和教学优化提供数据支持。研究表明, 所构建的改进 K-means 算法能够有效对学生进行分类并建立群体特征, 有利于教学质量与育人质量提升。

关键词: 人工智能; K-means 算法; 聚类分析

DOI: 10.65976/3105-4838.2026.02.004

引言

(一) 研究背景

在信息技术和人工智能快速发展的背景下, 教育正经历深刻变革。《中国教育现代化 2035》提出要推动人才培养模式改革, 实现规模化与个性化的结合。教育部也强调要深化教育评价改革, 利用大数据和人工智能构建更科学的学生发展评价体系, 推动教学决策从“分数驱动”向“过程数据驱动”转变。

随着教育大数据的深入应用, 学生学习过程数据日益丰富, 但如何挖掘数据背后的信息、揭示学生真实状态与需求, 仍是提升教育评价科学性的关键问题。

聚类分析作为无监督学习方法, 能有效识别学生群体特征与行为模式, 为精准育人提供依据。本文通过聚类算法分析学生学习行为数据, 提出针对性教学建议, 响应国家精准教育政策, 助力个性化教学与学生全面发展。

(二) 国内外研究现状

随着智能算法的发展, 数据挖掘技术在教育教学中得到广泛应用。学者们通过关联规则^[1]、支持向量机^[2]、决策树和聚类算法等方法^[3-6], 构建学习成效预测模型, 用于学业预警和教学干预。然而, 现有研究多基于单一或局部数据, 未能有效融合多维数据以揭示学生内在需求, 难以支撑精准教育策略的制定。

为实现精准教育支持, 需从学生群体特征出发进行分析。K-means 等聚类方法被广泛用于识别不同学生群体, 并据此调整教学内容与方式^[7-11]。尽管相关研究取得一定进展, 但聚类数目和初始中心选择对结果影响较大, 如何优化算法、提升分类准确性, 仍是推动精准干预和个性化教育的关键问题。

鉴于此, 本文采用改进的 K-means 聚类算法, 构建一个融合学生学习过程数据与学习成效数据的聚类模型。对在校学生多维度学习过程数据和学习成效数据进行聚类分析, 识别出具有相似学习行为和学习效果的学生群体, 从而揭示不同学生群体在学习过程中的共性与差异。基于聚类结果, 进一步提出针对性的教育教学改进策略, 通过数据驱动的方式为教师提供科学决策依据, 助力实现因材施教。

一、方法建立

(一) K-means 聚类算法基本原理

K-means 是一种迭代聚类算法, 旨在最小化簇内平方误差, 提升簇内样本相似性^[8]。给定 n 个样本的数据集, 每个样本有 m 个属性, 算法通过计算样本间相似性, 将它们分配到 K 个簇中, 每个样本仅属于相似性最高的簇。传统 K-means 随机选择 K 个点作为初始中心, 迭代分配样本并更新中心, 直至中心不再变化或达到最大迭代次数。

其中, 重新计算每个类簇中心时, 将同一簇中所有采样点的平均值作为参考点, 即簇的中心。公式表达为^[12]:

$$C_k = \frac{\sum_{X_n \in N_k} X_n}{|N_k|} \quad (1)$$

式中, C_k 为第 k 个聚类中心, X_n 表示第 k 个簇中第 n 个样本对象, $|N_k|$ 为第 k 个簇中样本个数。

样本与聚类中心的相似性用欧几里得距离计算^[12]:

$$Dist(S_n, A_k) = \sqrt{\sum_{m=1}^m (S_{n,m} - A_{k,m})^2} \quad (2)$$

基金项目: 重庆轻工职业学院教改项目 (QGZYJGYB202501)。

式中, S_n 为第 n 个样本, A_k 为第 k 个聚类中心, $S_{n,m}$ 表示第 n 个样本的第 m 个属性, $A_{k,m}$ 表示第 k 个聚类中心的第 m 个属性。

显然, 在传统 K-means 聚类算法中, 想获得一个聚类结果, 需要将聚类数目 K 从小到大依次尝试, 不断迭代寻找新的类簇中心点。值得指出的是, 由于达到最大迭代次数, 最终的聚类结果可能不是样本数据整体最优解, 而仅是局部最优解。因此针对传统 K-means 聚类算法迭代寻找类簇中心点的问题, 本文结合肘部算法, 运用“拐点”的数学本质, 通过量化方式确定 K-means 最佳聚类数目 K , 减少遍历次数实现整体最优解, 提高聚类结果的稳定性和准确性, 同时降低运用肘部算法确定 K-means 聚类数目时的主观性。以智能赋能指导教师根据不同的学生群体提出针对性的教育教学改进策略, 提升教师教育教学水平。

(二) 改进的 K-means 聚类算法

为了合理确定 K-means 聚类的最佳聚类数目 K , 结合肘部法则基本思想, 通过寻找总误差平方和 (Sum of Squared Errors, SSE) 下降速率显著减缓的“拐点”来确定最佳聚类数目 K 。在聚类运算中 SSE 作为评估簇紧密度的指标, 表示各数据点与其所属簇中心距离的平方和。理论上, 较小的 SSE 意味着簇内数据样本相似度更高, 即簇内样本间更接近。

总误差平方和 (SSE) 可用公式 (3) 表达:

$$SSE = \sum_{i=1}^k \sum_{x_j \in C_i} \|x_j - \mu_i\|^2 \quad (3)$$

式中, k 为聚类数目, C_i 为第 i 个聚类中的所有数据样本, x_j 为第 j 个数据样本, μ_i 为第 i 个聚类的中心点, $\|x_j - \mu_i\|^2$ 表示样本 x_j 到所属聚类中心 μ_i 的欧几里得距离的平方。

在拐点之前, 每增加一个聚类数目都会使 SSE 显著降低, 表明聚类结构可以被有效划分。在拐点之后, 聚类数目进一步增加不会使总误差平方和得到明显改善, 反而会加大模型复杂性并降低聚类 SC 轮廓系数。因此寻找“拐点”确定合适的聚类数目 K 是达到最佳聚类效果的关键。

具体方法步骤如图 1 所示。

二、学生学习成效与个体状态聚类分析

为了验证所建方法的可行性, 本文收集了某高校理工科专业 7 个班 209 个学生数据, 进行改进的 K-means 聚类算法验证分析, 并结合聚类结果对学生群体进行特征分类, 探究不同类型学生内在需求, 为不同的学生群体提出有针对性的教育教学改进措施。

(一) 肘部算法运算结果

为了多维度考量学生学习成效与各因素之间的关系, 本文收集了某高校理工科专业 7 个班 209 个学生数据, 内容包括“理工学科成绩”“人文学科成绩”“心理评估”“学习动机指针”“出勤率”“体育健康程度”6 个方面, 评判学生心理状态、学习习惯、内驱力对学习成效的影响。

根据所建立的方法, 在进行 K-means 聚类分析前, 需先采用肘部算法确定最佳聚类数目 K 。

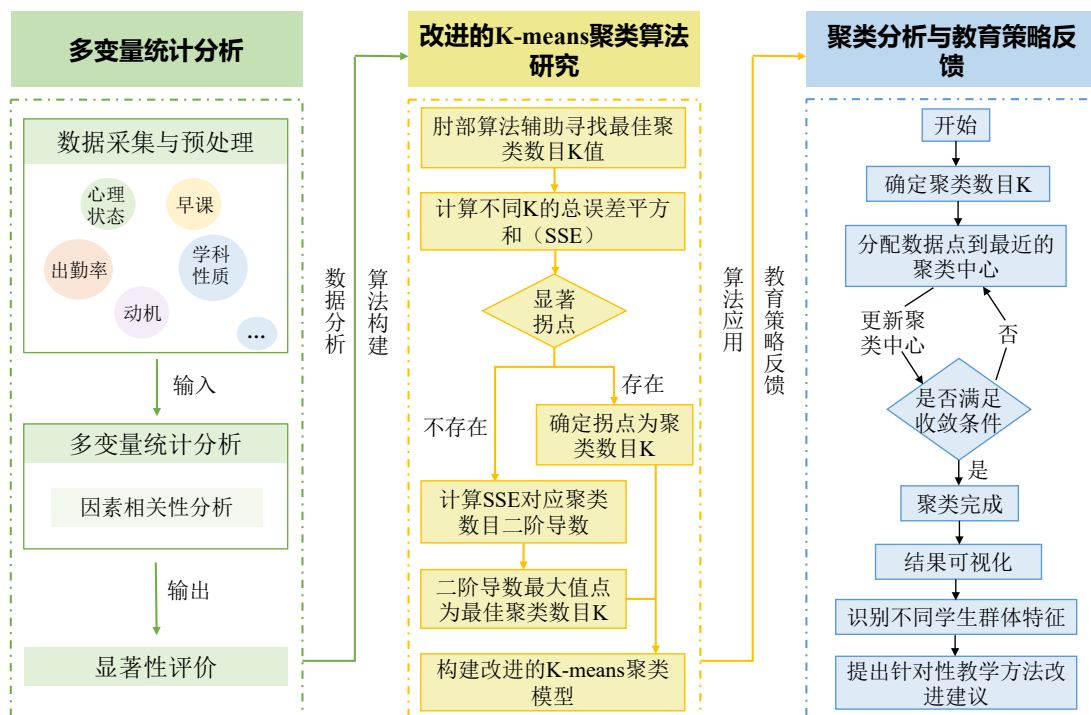


图 1 方法步骤流程图

图2为总误差平方和与聚类数目关系图，横坐标表示聚类数目K，纵坐标表示K所对应的总误差平方和(SSE)。从图2可以看出，随着K值的增加，SSE逐渐减少。这是因为当聚类数目增多时，每个簇内的数据集规模会缩小，使得每个样本能够选择更接近的聚类中心，因此，样本点到聚类中心的平均距离减小，从而导致SSE降低，注意到，当K=4时，SSE下降速率变缓，形成类似于“肘部”的拐点，根据肘部算法结果，确定最佳聚类数目K=4。

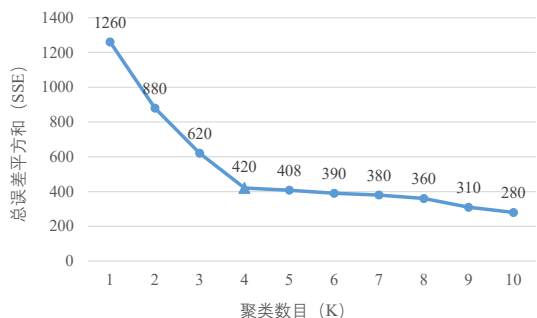


图2 肘部法则确定最佳聚类数目

(二) 改进的K-means 聚类算法运算结果

由于六维数据难以可视化，也不利于直观分析。为了可视化聚类结果，本文结合主成分分析(Principal Component Analysis, PCA)，将六种基础数据(理工学科成绩、人文学科成绩、心理评估、学习动机指针、考勤率、体育健康程度)降至两个维度，降低数据冗余，同时尽可能多的保留有效信息。主成分1和主成分2是通过PCA从原始数据中提取的两个主要特征。这两个主成分为原始数据中方差最大的两个线性组合，二者相互正交，即它们之间没有相关性。可视化聚类结果散点图如图3所示。

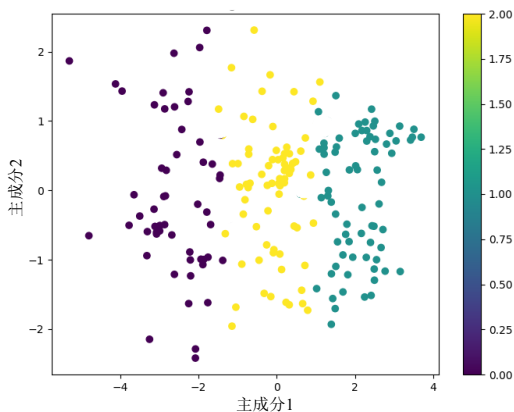


图3 聚类结果散点图

如图3所示，聚类结果被划分为3类，三类点形成三个明显的“云团”，云团中心距离较远，同类点聚集紧密。这说明在聚类数目K=3时形成了良好的聚类效果。通过改进的K-means 聚类算法将六组特征数据样本聚类于三种标签类型中，构建分类标签如表1所示。

(三) 基于聚类结果提出针对性教育教学改进建议
在进行针对性教育教学改进建议之前，需要对各类簇进行特征分析，明确聚类群体内在需求。

1. 高绩优型

高学习表现型学生在传统学业成就上表现最为突出，文理课程得分均显著高于其他群体，但心理状态或体育健康相对一般。这表明高绩优型学生本身具有扎实知识基础、高学习天赋或投入了大量时间和精力，而忽视了其他方面。暗示其精力可能高度集中于学习领域，在综合健康方面存在较大提升空间。

2. 综合素养型

综合素养型学生是“全面发展”的典范，没有明显的短板，在各个维度上均表现出良好水平。这类学生最大优势不在于某个维度上的突出表现，而在于多维度上的协调发展。他们像“六边形战士”，虽然没有一个维度达到极致，但整体结构最为稳固。代表了教育目标中“德智体美劳全面发展”的典范，适应性强，后续发展潜力持久。

3. 发展提升型

发展提升型学生在当前传统学业评价中不占优势，但其数据模式揭示出他们拥有独特的、未被充分激发的潜力。发展提升型学生心理状态和学习动机均较好，学习动机可能并非指向传统学业竞争，而是与实践活动、社交认同或身体运动相关。他们可能具备较强的动手能力或高情商，只是尚未与学术学习有效结合。这类学生在“低学业表现”表象之下，是未被传统教学活动激发的多元智能。教育干预的关键在于找到其优势领域(如运动、实践、社交)与学术学习之间的连接点，进行动机迁移。

通过上述分析，可以看到每种聚类的特点，这些特点可以帮助教师理解学生群体差异，并为不同学生群体提供更有针对性的教育支持措施。

对于高绩优型学生，鼓励其保持学习成效优势的同时，着重引导其认识全面发展的重要性。关注社交技能、情绪管理、心理调试等能力的培养。鼓励通过

表1 聚类标签类型及指标

标签名称	标签指标
高绩优型	强调学业成就的突出表现，但心理状态或学习动机不强
综合素养型	综合素质均衡，各方面平衡发展
发展提升型	学习成效表现不足，需要通过支持来释放潜力

探索与校内学习科目无关的兴趣爱好,通过非学科兴趣丰富个人生活体验,平衡生活重心。同时,可以鼓励其担任学生干部、参与社团活动等方式,提升团队协作能力和责任感。

对于综合素养型学生,应鼓励其继续保持全面协调发展,并在此基础上进一步加强学习深度与心理韧性建设。教师可在教学实践中设计融合学科探索课题或实践项目,提高面对实际问题的应对能力;通过组织辩论、案例分析等活动,锻炼其批判思维与复杂问题应对能力。此外,学校可提供系统的职业规划指导,帮助学生明确发展方向,将当前综合优势转化为可持续的成长路径。

对于发展提升型学生,应该重点激发其学习潜力并维持积极心理状态。可借助他们在运动、实践或社交方面的优势,设计与其兴趣相连接的学习任务,确保学生在感兴趣的领域得到充分发展,并以此建立学业的信心桥梁。另外,可以通过制定个性化学习计划,适当学习成效认可和过程激励,提高这部分学生的学习追求和自我效能感,帮助学生在受支持的环境中稳步提升,实现潜力有效转化。

三、结论与展望

本文聚焦于高校学习评价流于表面、维度单一,难以发现学生特征和内在需求,从而难以为“因材施教”提供有效教育教学策略的现实问题,提出一种改进的K-means聚类算法,构建聚焦学生学习过程数据与学习成效数据的聚类模型,实现学生学习特征精准分类。并采用实际数据对所建立的方法进行了算法验证,验证结果表明所构建的聚类方法能够实现群体精准分类,帮助教师识别不同群体的学习模式与潜在需求,为开展精准教育教学干预提供数据支持。在此基础上,根据聚类结果提出有针对性的教育教学改进策略,提升教育评价科学性、实现精准育人。

需要指出的是,本文为提出的改进的K-means聚类算法为学生群体特征识别与教学干预提供了技术支撑和实证探索,但由于能采用的数据源有限,在基础

数据质量与维度上,后续可加强“数字校园”建设获取更多在校学生行为数据,为进一步挖掘学生内在需求提供更开阔的数据支撑。

参考文献:

- [1] 薛新力,郑建盛,林志萍.家长心理控制源与子女学习成绩的相关研究[J].中国行为医学科学,2002(02):89-91.
- [2] Anupam Khan,Soumya K.Ghosh.Data mining based analysis to explore the effect of teaching on student performance[J].Education and Information Technologies,2018,23(4):1677-1697.
- [3] 刘博鹏,樊铁成,杨红.基于数据挖掘技术的学生成绩预警应用研究[J].四川大学学报(自然科学版),2019,56(02):267-272.
- [4] Bezerra L N M,Silva M T.Educational data mining applied to a massive course[J].International Journal of Distance Education Technologies(IJDET),2020,18(04):17-30.
- [5] 陈洪涛.基于K-means算法的学生行为数据与学习成绩关联性分析[J].中国科技信息,2024(23):86-88.
- [6] 郑安琪,王宇琪,郝川艳.基于深度学习的在线学习成绩预测研究[J].计算机时代,2021,(12):69-72+75.
- [7] 娄德涵.高校学生行为习惯与成绩关联性分析研究[D].西安:西京学院,2022.
- [8] 申丹丹.基于R语言的主成分分析与聚类分析在绩效评价中的应用[J].科技资讯,2024,22(21):253-256.
- [9] 张梁,杨立波,张小勇,等.基于优化K-means算法的高校成绩聚类分析研究[J].太原学院学报(自然科学版),2024,42(02):79-84.
- [10] 李森林,张肇龙.成人学习特点分析与继续教育教学改革改进[J].成人教育,2024,44(12):9-16.
- [11] 陆觉民,马国栋,郑宇.基于数据挖掘技术的图书馆流通数据的关联分析[J].现代情报,2009,29(09):108-110.
- [12] 王森,刘琛,邢帅杰.K-means聚类算法研究综述[J].华东交通大学学报,2022,39(05):119-126.