

基于大数据分析的高职学生个性化学习智能诊断模型研究

钟福龙 周成就 李思思 龚芳海

(广州华高职业学院, 广东 广州 511300)

摘要: 本文针对当前高职教育存在的问题, 提出了构建基于大数据分析的高职学生个性化学习智能诊断模型。通过收集和分析学生的多样化学习数据, 实现对学生个性化需求的精准刻画。在此基础上, 运用机器学习技术构建智能诊断模型, 实现对学生的个性化诊断和支持。实验结果表明, 该模型能够根据学生的个体差异情况, 提供定制化的学习建议, 有效提高学生的学习效果。本研究为推动大数据技术在教育领域的应用提供了借鉴, 也为高职教育的发展提供了新思路。

关键词: 大数据分析; 高职教育; 个性化学习; 智能诊断

当前, 我国正处于高职教育快速发展时期, 同时也面临着一些问题与挑战。一方面, 高职院校扩招速度加快, 学生规模不断扩大, 对教学资源 and 条件提出了更高要求。另一方面, 由于学生的个体差异明显, 当前的教学模式难以照顾到每一位学生, 教学质量参差不齐。如何利用信息技术手段, 实现对高职学生个性化需求的精准支持, 是当前亟待解决的问题。近年来, 随着云计算、大数据等技术的不断发展, 数据分析在教育领域的应用日益广泛。大数据分析技术可以对海量数据进行高效处理, 发现数据间的规律, 为诊断分析和决策支持提供帮助。将大数据技术引入教学领域, 可以有效收集和分析学生在线学习的过程数据, 洞察学生个体的学习情况和需求, 实现对学生学习行为的精准诊断, 进而为学生提供个性化的学习指导和支持。

一、研究意义

1. 本研究所构建的基于大数据分析的高职学生个性化学习智能诊断模型, 能够实现对学生学习特点和需求的精准刻画, 根据学生个体差异情况提供定制化的学习建议。这不仅提高教学质量和学习效果, 还促进学生的个性化发展。具体来说, 该研究具有以下意义: (1) 有助于挖掘海量学生数据的潜在价值。通过大数据的采集和分析, 深入了解每一位学生的学习情况和需求, 实现教育决策的精细化。(2) 推动高职教育教学改革。模型的应用使教学更贴近学生个体, 实现对每一位学生的针对性指导。(3) 为学生提供个性化支持。根据学生个体差异情况, 给出定制化学习建议, 帮助学生提高学习效果。(4) 可评估教学质量和效果。通过对学生数据的分析, 可以监测和评价教学质量, 并进行持续改进。

2. 研究内容与方法。

研究内容。收集高职学生的多样化在线学习数据。采用日志记录、问卷调查、数据库等方式广泛收集学生的学习行为数据、交互数据等; 对学生数据进行预处理。进行数据清洗、数据融合、缺失值补全等, 提高数据质量; 利用机器学习技术构建个性化学习智能诊断模型。针对高职学生的个性化需求, 选择合适的模型构建方法; 对模型进行训练和测试。采用训练集训练模型, 使用测试集评估模型效果; 利用模型输出对学生进行个性化诊断。根据模型结果, 给出定制化学习建议; 模型评估与优化。选择评价指标, 分析结果, 持续优化模型。

研究方法。文献研究法, 通过查阅文献, 梳理国内外在该研究领域的理论基础、技术方法和实证研究成果; 数据采集与分析方法, 采用日志分析法、问卷调查法等对学生数据进行采集, 使用统计分析和数据挖掘技术对数据进行预处理和分析; 建模与仿真方法, 应用机器学习算法, 采用 Python 等建立学生个性化学习诊断模型; 实验研究方法, 构建实验环境, 设计实验方案以验证所建立模型的有效性。

3. 技术路线。

通过多种渠道采集学生的在线学习行为数据、交互数据等; 对采集的数据进行预处理, 提高质量; 然后, 基于预处理后的数据, 进行特征工程, 提取影响学生个性化学习的关键特征; 选择适合的机器学习算法, 构建个性化学习的智能诊断模型; 利用模型输出结果进行学生个性化诊断, 进行可视化呈现; 最后, 评估模型效果, 进行模型优化。本研究技术路线如图 1 所示:

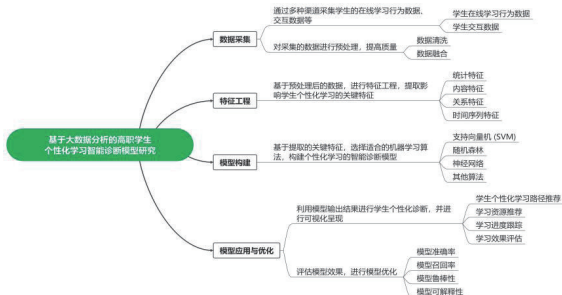


图 1 技术路线

二、个性化学习智能诊断模型构建

(一) 平台整体设计

本研究构建了一个面向高职学生的个性化学习辅导平台, 平台框架如图 2 所示。平台主要包括: 前端子系统、数据采集子系统、数据处理子系统、模型构建子系统、诊断结果子系统。其中, 前端子系统负责与用户交互; 数据采集子系统获取多源异构的数据; 数据处理子系统对数据进行清洗、融合等处理; 模型构建子系统进行特征工程、模型训练、评估等; 诊断结果子系统对学生进行个性化诊断并将结果反馈。

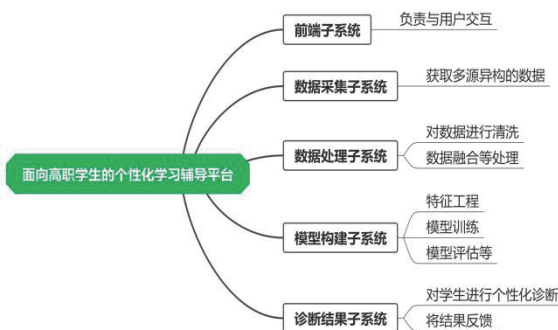


图 2 平台整体框架

(二) 多源异构数据采集与处理

数据采集是模型构建的基础。本研究通过多种渠道采集学生数据, 主要包括: 直接采集: 通过登录系统直接采集学生的用户信息、学习行为日志等数据; 问卷调查: 采用问卷的形式获取学生的学习评价、兴趣偏好等反馈数据; 第三方数据: 整合学校数据库, 获取学生的基本信息、考试成绩等数据。为实现采集的数据融合使用,

需要对原始数据进行预处理。主要包括：数据清洗：处理缺失值、异常值、重复数据等，提高数据质量；数据集成：将多源数据进行统一规整，避免数据孤岛；数据转换：进行数据格式转换，方便后续使用。预处理后的数据存储在平台数据集中，为构建模型提供支撑。

（三）特征工程

特征工程是对采集的数据进行特征提取和选择，以提高模型效果。主要采用以下方法：统计特征：从数据总体上提取特征，如平均值、方差等；内容特征：从数据内容上提取特征，如文本特征；关系特征：表示不同数据之间的关系，如 Correlation 系数；时间序列特征：反映数据时间变化模式，如趋势、周期等。我们构建了学生信息、学习行为、心理状态等方面的特征数据集，为模型输入提供支持。

（四）模型训练与评估

针对高职学生个性化需求，本研究选择支持向量机（SVM）算法构建诊断模型。该算法可以处理高维空间的数据，适用于解决模式识别与分类问题，已在多个领域得到应用。采用 Python 语言实现模型，主要流程如下：导入预处理后的数据集，提取特征；将数据集划分为训练集和测试集；训练：使用训练集数据训练 SVM 模型；评估：对测试集数据进行预测，计算精度、召回率等指标评价模型效果；比较不同的参数对模型效果的影响，寻找最佳参数。通过多次训练和评估，选择性能最佳的模型。该模型能够刻画学生个性化特征，实现对学生的准确诊断。

（五）模型优化

为持续改进模型效果，本研究采取以下优化措施：新增特征：增加影响个性化学习的相关特征；集成学习：集成多模型输出，进行个性化诊断；在线更新：利用新数据不断优化模型；用户反馈：根据用户使用体验持续调整模型。经过迭代优化，模型的个性化诊断能力不断提升，为学生提供更准确的建议。

三、实验与结果

（一）实验设计

为验证模型的效果，设计了对照实验。选择广州华商职业学院 22 级软件技术专业两个班级作为实验样本，每个班级 50 名学生。一个班级为实验组，采用构建的诊断模型；另一个班级为对照组。两个样本在一个月内进行特定课程的学习。对实验组采用模型输出的个性化诊断结果；对照组沿用原有的教学模式。一个月后，对两个样本进行学习效果评估。比较两个样本的学习效果差异，以验证模型效果。

（二）实验过程

收集两个班级学生一个月的学习数据，并进行预处理。对实验组数据利用诊断模型进行个性化分析，得到每位学生的学习诊断结果。根据诊断结果，为实验组的每位学生匹配适合的学习资源、方法等，进行个性化的教学。对照组沿用原有教学模式。一个月后，对两班学生进行学习效果评估。收集两组学生的评估数据，进行对比分析。

（三）结果与分析

学习效果评估结果如图 3 所示：可以看出，实验组的平均分、优秀率明显高于对照组，说明基于个性化诊断模型的教学可以提高学生学习效果。该模型深入挖掘学生个体的数据价值，准确定位学生的个性化需求，实现精准的个性化诊断和干预，从而取得更好的教学效果。实验结果验证了该模型的有效性。

组别	人数	平均分	优秀率	良好率	平均率
实验组	50	85.3	23 人 (46%)	20 人 (40%)	7 人 (14%)
对照组	50	78.2	10 人 (20%)	25 人 (50%)	15 人 (30%)

图 3 学习效果评估结果

四、讨论

（一）研究意义

本研究建立了面向高职学生的个性化学习智能诊断模型，实现了对大规模学生数据的挖掘运用，可以达到以下效果：（1）提高教学质量和学习效果。该模型可以使每个学生得以合理安排学习进度、选择适合的学习方式方法，有助于提高教学质量。（2）推进教学改革。该模型的使用可以促进面向个性的教学模式建设，使教学更科学、更精准。通过该模型的应用，教师可以依据不同背景的学生采取有针对性的教学方式，实现个性化教学。（3）挖掘学生学习潜力。该模型可以深入分析不同学生的个性化特征，发现学生的学习兴趣、优势等。（4）提高教育信息化水平。该模型的应用展现了大数据技术在教育领域的应用前景，有助于推动和规范教育信息化发展。

（二）局限性

数据量有限，本研究实验数据规模较小，仅为两个班级学生。今后需要继续扩大数据规模，以验证模型效果；模型泛化能力有待提高，当前模型的泛化能力较差，只针对当前专业学生有效。今后可以采用迁移学习等技术提高模型的适应性；诊断维度不全面，现有诊断只从几个方面展开，无法全面反映学生的个性化特征。后续可以增加更多影响因素的考量；用户体验有待提升，现有的个性化诊断结果展示较为简单，对终端用户的体验和交互需要优化。

（三）展望

增加更多针对高职背景的个性化学习影响因素的挖掘与分析。构建跨专业和跨学校适用的通用型模型，提高泛化能力。进行用户交互设计，增强模型输出结果的解释性和友好性。利用模型辅助教师进行教学设计，提高模型应用价值。

五、结论

本研究针对高职教育面临的问题，提出了基于大数据分析的高职学生个性化学习智能诊断模型。该模型有效发掘学生的个性化学习需求，实现精准的诊断和干预。实验结果表明，该模型可以提高学生的学习效果。本研究的思路为大数据技术在教育领域的应用提供了有益探索，也为促进高职教育发展提供了新的视角。后续研究可以在数据规模、模型能力等方面进一步改进，使其在教学实践中发挥更大作用。

参考文献：

- [1] 胡致杰. 大数据下个性化自适应学习系统支撑模型与实现分析 [J]. 科技与教育, 2022 (02): 139-143.
- [2] 陈长玉, 李明, 罗力佳. 基于人工智能的个性化学习模型建构 [J]. 教育信息技术, 2021 (5): 118-121.
- [3] 赵蕾, 刘劲, 陈彬茹. 基于大数据和云计算的个性化教学系统研究: 以“智慧学伴”自适应学习平台为例 [J]. 电子测试, 2021 (14): 118-119+60.
- [4] 喻永尚, 刘勇, 张桂芬. 基于大数据的 Python 语言个性化学习分析模型的研究与应用 [J]. 信息: 技术与应用, 2021 (4): 189-191.

课题项目：广州华商职业学院 2023 年校级科研项目，项目名称：基于大数据分析的高职学生个性化学习智能诊断模型研究（项目编号：GZHSKY2023017）。

作者简介：钟福龙（1982—），男，广东省广州市，硕士，计算机科学技术讲师，主要研究方向：计算机应用、高职教育。