·教育研究·

临床医学专业课程学习者画像模型构建

刘蕾[1]　王蕾[2]　王晨[3]　王云[3]　高清[4]
[1]首都医科大学附属北京同仁医院人力资源处，北京　100730；[2]中国医学科学院医学信息研究所信息技术部，北京　100020；[3]首都医科大学附属北京朝阳医院麻醉科，北京　100020；[4]首都医科大学附属北京朝阳医院教育处，北京　100020
通信作者：刘蕾，Email：l.liu_trhos@foxmail.com

【摘要】 目的　以“超声在麻醉、疼痛和重症医学中的应用”课程为例，构建临床医学专业课程学习者画像模型。**方法**　梳理临床医学专业课程学习者画像模型构建框架及流程，对多来源学习者数据进行收集及预处理。利用Python 3.9编程语言进行统计分析、自然语言处理、聚类分析，并以可视化技术呈现，构建机器学习预测模型并评估模型预测效能。**结果**　从学习背景、兴趣偏好和行为效果3个维度构建临床医学专业课程学习者画像模型。学习背景画像揭示了学习者基本信息、认知基础和学习动机。兴趣偏好画像分析了学习目的与其他选课信息，根据内容关注程度、学习效果影响因素识别出3类不同的学习者群体。行为效果画像通过构建4种机器学习算法预测模型实现了对课程考试成绩的分类预测，结果显示朴素贝叶斯算法效果最佳，准确率为0.80、F1分数为0.79，受试者工作特征（receiver operating characteristic，ROC）曲线下面积（area under the curve，AUC）为0.79（P=0.035），与其他算法差异有统计学意义。**结论**　本研究构建了临床医学专业课程学习者画像模型并开展实证研究，画像结果对教学内容、教学方式与教学团队、学习效果预测提供了指导建议。

【关键词】 学习者画像；　临床医学；　聚类分析；　机器学习；　可视化
【中图分类号】 R-4
基金项目：首都医科大学教育教学改革研究课题（2022JYY193）
DOI: 10.3760/cma.j.cn116021-20230205-01442

Construction of learner profile models for clinical medical courses

Liu Lei[1], Wang Lei[2], Wang Chen[3], Wang Yun[3], Gao Qing[4]
[1]Department of Human Resources, Beijing Tongren Hospital, Capital Medical University, Beijing 100730, China; [2]Division of Information Technology, Institute of Medical Information, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100020, China; [3]Department of Anesthesiology, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China; [4]Department of Education, Beijing Chaoyang Hospital, Capital Medical University, Beijing 100020, China
Corresponding author: Liu Lei, Email: l.liu_trhos@foxmail.com

【Abstract】 Objective　To construct a learner profile model for clinical medical courses with the course "Application of ultrasound in anesthesia, pain and critical care medicine" as an example. **Methods**　The framework and procedures were established for constructing the learner profile model for clinical medicine courses, and learner data from multiple sources were collected and pretreated. Python 3.9 programming language was used to perform statistical analysis, natural language processing, and cluster analysis, and the results were presented using visualization techniques. Machine learning prediction models were established and evaluated in terms of prediction performance. **Results**　The learner profile model for clinical medical courses was established from the three dimensions of learning background, learning interests and preferences, and learning behaviors and effects. The learning background profile revealed the basic information, cognitive foundation, and learning motivation of learning. The interests and preferences profile analyzed the learning objectives and other selective courses of learners, and three different groups of learners were identified based on content attention and influencing factors for learning effects. The behaviors and effects profile achieved the classified prediction of course examination scores by constructing four machine learning algorithms, and the results showed that the naive Bayes algorithm had the best effect, with an accuracy of 0.81, an F1 score of 0.79, an area under the receiver operating characteristic curve of 0.79, and a

P value of 0.035, indicating significant differences between the naive Bayes algorithm and other algorithms. **Conclusion** This study constructs a learner profile model for clinical medical courses and conducted an empirical study, and the results of this model provides guidance and suggestions for teaching contents, teaching methods, teaching team, and learning effect prediction.

【**Key words**】 Learner profile; Clinical medicine; Cluster analysis; Machine learning; Visualization

Fund program: Education and Teaching Reform Research Project of Capital Medical University (2022JYY193)

DOI: 10.3760/cma.j.cn116021-20230205-01442

《中国教育现代化2035》指出，鼓励基于大数据分析，加强对信息时代学习者认知和学习行为规律的研究[1]。用户画像技术应用于高等教育领域，能够发现教育用户特征和教育活动特点、发掘教学规律并预测发展趋势[2]。超声引导神经阻滞是麻醉与疼痛治疗的一项重要技术[3]。首都医科大学附属某临床医学院于2018年率先开设“超声在麻醉、疼痛和重症医学中的应用”课程。此课程通过理论讲解和实操演示，培养学习者超声引导下区域麻醉、疼痛治疗和评估危重病人状态的专业技术，提升了学习者在临床实践中应用便携式超声设备解决复杂临床问题的思维和能力。通过4年来的实践探索，该课程教学体系与课程设计虽渐臻完备，但仍面临教学内容繁多、涉及多学科交叉、理论与实践转化等问题。加之该类课程普遍存在学习者参与度不高、操作协调能力弱等学习困难[3-4]。据此，借助教育教学中的新技术和新方法，结合学生专业特点和课程特色，有针对性地提出问题对策，是当前临床医学类课程教学亟待研讨的问题。在此背景下，将用户画像技术应用于临床医学教育与专业课程教学，不失为一个有益尝试。本研究以“超声在麻醉、疼痛和重症医学中的应用”课程为例，构建临床医学专业课程学习者画像模型，深入了解其学习背景、兴趣偏好和行为效果，为授课团队调整教学内容、优化教学过程、提升教学质量提供参考建议。

1　研究现状概述

用户画像被定义为“基于用户真实数据的虚拟代表”[5]。已有学者积极探索用户画像在教育教学中的应用，研究主要涉及：教育领域用户画像算法与技术[6-9]、教学参与者画像构建[10-13]、用户画像在特定学科或管理场景中的应用等[14-16]。然而在医学教育领域，针对医学专业课程的学习者画像研究却较少涉及。本研究利用用户画像技术，全面了解临床医学专业课程学习者特征，进而提高难度较大的课程的教学质量和学习效果，高效推进课程发展建设。

2　研究设计与方法

2.1　研究设计

借鉴既有学习者画像模型构建相关研究成果[17-18]，结合课程实际情况和授课团队需求反馈，从学习者学习背景、兴趣偏好、行为效果3个维度构建临床医学专业课程学习者画像模型。流程包括数据来源与收集、数据处理与标签化、标签体系与画像建模、可视化展示与结果分析、画像输出与教学指导5个步骤。根据画像结果分析讨论，密切联系课程教学情况，从课程建设层面提出指导建议供教学团队参考。模型构建框架如图1所示。

2.2　数据来源与预处理

以参与2022年至2023学年度第一学期“超声在麻醉、疼痛和重症医学中的应用”课程学习者31人为研究对象。为实现学习者画像的相对动态性，在研究过程中对于学习者原始数据收集的时间范围覆盖课前、课中、课后3个阶段。通过首都医科大学研究生管理系统、问卷星“超声在麻醉、疼痛和重症医学中的应用”课程学习者调研（ID：192093254）和腾讯会议签到记录，获取画像标签数据（表1）。对原始数据进行清洗，处理冗余值、缺失值和异常值。所有研究对象对本研究均已知情同意。

2.3　学习者画像建模

本研究采用统计分析、自然语言处理、聚类分析及机器学习算法构建学习者学习背景画像、兴趣偏好画像和行为效果画像。

2.3.1　描述性统计分析

采用Python 3.9软件，对由“首都医科大学研究生管理系统”和“腾讯会议签到记录”导出的数据进行描述性统计。计数资料以百分比表示，计量资料以（均数 ± 标准差）表示。

2.3.2　自然语言处理

对原始数据为非结构化的自然语言进行文本挖掘。通过基于Python编程语言的Jieba库分词、自定义词典、去除停用词后，进行词频统计，获取用以可视化展示的词云标签。

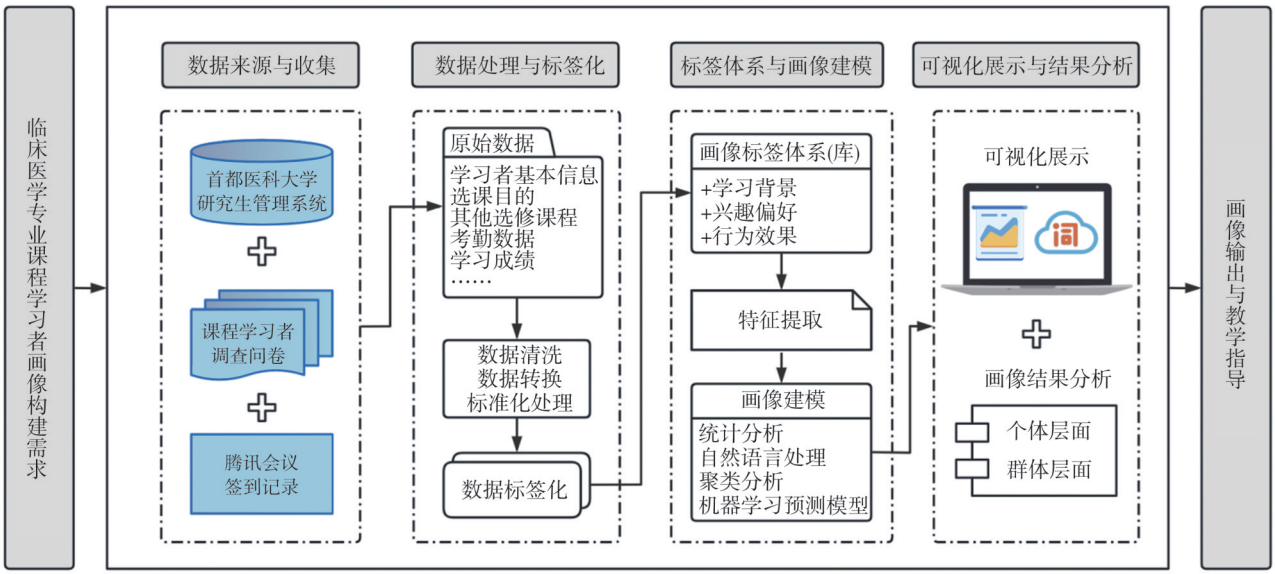


图 1　临床医学专业课程学习者画像模型构建框架

表 1　原始数据与画像标签

数据来源	数据属性	数据形式	画像标签	取值说明
首都医科大学研究生管理系统	学号	数值型	L0 固定属性	学习者学号
	姓名	结构化文本型		学习者姓名
	性别	分类型	L1 基本信息	男；女
	攻读学位	分类型		博士；硕士
	入学方式	分类型		统招；同等学力
	专业	分类型		学习者专业
	课程考试成绩	数值型	L2 课程成绩	学生笔试成绩，满分 100
课程学习者调查问卷	开课前课程了解程度	分类型	L3 认知基础	特别了解；大概了解；不了解
	选择课程的主要原因	分类型	L4 学习动机	认知兴趣；专业提升；职业发展；外部影响
	学习课程的主要目的	非结构化文本型	L5 学习目的	自然语言
	课程教学内容关注程度	数值型	L6 关注程度	对课程教学计划中的授课内容关注度赋分（1—基本不关注；2—关注一些；3—般关注；4—比较关注；5—非常关注）
	其他选修课程名称	非结构化文本型	L7 选课信息	自然语言
	影响学习效果的教师因素	分类型	L8 影响因素	教师教学态度；教师语言风格；教师相貌、年龄和气质；教师学历学识
	影响学习效果的课程因素	分类型		课程授课方式；课程考勤制度；课程考核方式；课堂互动交流；日常扩展阅读；课程管理模式
	是否使用信息化辅助学习工具	分类型	L9 资源利用	是；否
腾讯会议签到记录	课程出勤得分	数值型	L10 课程参与	2 分/次 × 10 次，满分 20

2.3.3　兴趣偏好聚类分析

聚类分析选择无监督学习 Kmeans 算法（K-均值聚类算法），识别不同学习者群体的兴趣偏好。对 L6 标签的数值型变量和 L8 标签的分类型变量分别进行标准化处理和独热编码，应用肘部法确定聚类簇的个数 K，并以轮廓系数值验证聚类模型效果。

2.3.4　学习效果预测模型

构建课程学习者考试成绩的分类预测模型。因研究实际可获取数据集体量较小，为避免过拟合现象，基于原始数据，使用 Synthetic Data Vault 模型的 Gaussian Copula 算法进行数据增强[19]。根据样本量计算方法[20]，生成具备建模标签属性的课程学习者合成数据 99 例，数据整体质量分数为 94.29%。将课程学习者的真实数据随机划分为训练集（70%）和测试集（30%），另将合成的课程学习者数据加入训练集，共同参与模型构建。

选择逻辑回归(logistic regression)、朴素贝叶斯(naive Bayes)、决策树(decision tree)和随机森林(random forest)4种复杂度较低的机器学习算法,利用训练集构建预测模型并在测试集上检验。为获得最优的训练数据特征,在特征工程时采用独热编码处理多分类型变量。根据准确率、F1-score和受试者工作特征(receiver operating characteristic,ROC)曲线下面积(area under the curve,AUC),综合评价各算法构建模型的效能。AUC比较采用t检验,检验水准α=0.05。

2.4 学习者画像可视化

本研究使用matplotlib、wordcloud等基于Python编程语言的第三方包库实现可视化,对兴趣偏好画像的各标签值从学习者个体层面以词云形式展示。

3 结果

3.1 学习背景画像

选取学习者L1、L3和L4标签数据进行描述性统计分析,形成学习背景画像(表2)。在31名课程学习者中,女性71人(71.0%)、攻读硕士学位16人(51.6%)、以同等学力方式入学者21人(67.7%)。参与课程学习者的普遍认知基础良好,24名(77.5%)学习者均在某种程度上对课程有所了解。因实证临床专业课程涉及麻醉与超声交叉学科,以麻醉学和超声医学为专业的学习者占比67.7%(21/31),51.6%(16/31)的学习者学习动机为专业提升。

表2 “超声在麻醉、疼痛和重症医学中的应用”课程学习者背景信息

画像标签	标签值	频率	百分比(%)
性别	女	22	71.0
	男	9	29.0
攻读学位	硕士	16	51.6
	博士	15	48.4
入学方式	同等学力	21	67.7
	统招	10	32.3
专业	麻醉学	17	54.8
	超声医学	4	12.9
	其他专业	10	32.3
认知基础	大概了解	22	71.0
	不了解	7	22.6
	特别了解	2	6.5
学习动机	专业提升	16	51.6
	认知兴趣	9	29.0
	职业发展	4	12.9
	外部影响	2	6.5

3.2 兴趣偏好画像

为了掌握课程学习者的潜在需求和兴趣,探究其学习课程的主要目的和其他选课信息。对L5标签数据进行自然语言处理,按词频统计排名前5位的分别为“神经阻滞”“疼痛治疗”“超声诊断”“实际应用”和“临床”,说明学习者的主要学习目的为获得这些方面的提升。同时,对经过课程名称标准化处理后的L7标签数据统计词频,排名前3位的分别为“超声医学”“急诊医学”与“疼痛学”。由此可见,学习者的其他选课信息也与所学专业和本课程内容密切相关。

使用Kmeans算法对L6和L8标签数据进行聚类分析,挖掘课程学习者的专业知识关注度及学习效果影响因素。选择肘部法确定K值为3,轮廓系数为0.37,说明模型效果尚可。模型识别出3类不同偏好的学习者,其中,第一类学习者18人(58.0%),对课程内容关注度较高,认为教师教学态度、课程授课方式是影响其学习效果的主要因素;第二类学习者5人(16.0%),对课程教学内容关注度不高,认为教师的教学态度、语言风格、学历学识和课程授课方式、考勤制度影响着其学习效果;第三类学习者8人(26.0%),对各教学内容关注度尚可,认为除第二类学习者选择的因素外,课堂互动交流也对其学习效果产生影响。根据聚类结果,对各类别学习者从个体层面进行可视化展示。以第二类学习者为例,如图2所示。

3.3 行为效果画像

该学期参与课程学习者的平均出勤得分为(18.7 ± 3.0);考试成绩为(90.4 ± 3.5)分,根据平均值划分为成绩优秀者(≥90分)18人(58.1%)和成绩普通者(<90分)13人(41.9%)。参考关于学生成绩影响因素的研究成果[11,21-22],研究选取L1(入学方式、攻读学位和认知基础)、L6、L9和L10标签数据作为预测因子,构建L2标签的分类预测模型。对逻辑回归、朴素贝叶斯、决策树和随机森林4种机器学习算法构建的预测模型在测试集上的表现进行评估,其中朴素贝叶斯在准确率(0.81)和F1分数(0.79)2个指标上的值均相对较高。图3显示,朴素贝叶斯模型的AUC(0.79)值最高,t检验结果显示该模型的AUC与其他三者的差异具有统计学意义(P<0.05),且AUC在0.7~0.9时有一定准确性,提示该模型预测效果较好,故可采用该方法实现学习效果标签的有效预测。



图2　第二类学习者兴趣偏好个体画像

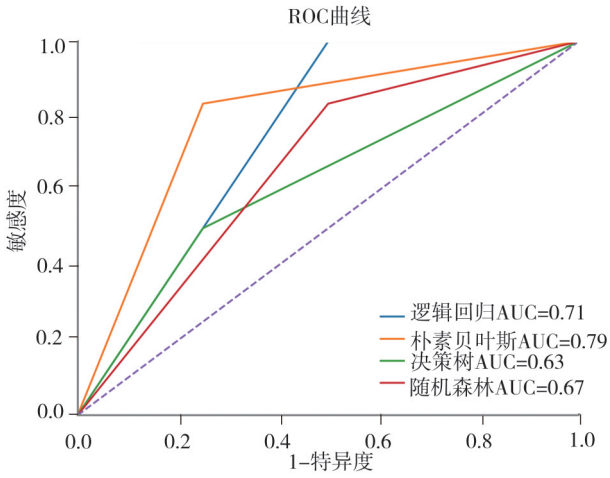


图3　课程学习者考试成绩预测模型的ROC曲线

4　讨论

4.1　学习者画像对实证课程建设的指导建议

课程教学内容方面，学习背景画像结果可帮助教学团队在开课前全面、快速了解授课对象背景信息，及时调整教学内容。本学期以麻醉学和超声医学专业为主的学习者达67.7%（21/31），且51.6%（16/31）的学习者的学习动机为专业提升、77.4%（24/31）的学习者开课前对课程有所了解。因此可实现分层教学，针对高水平组增加课程专业知识的深度和广度，避免重复性教学。此外，学习者参与课程的主要期待集中在“神经阻滞”“疼痛治疗”和“超声诊断”，且其他选修课程也与这些内容紧密相关，说明较多学习者希望通过课程获得针对性提升。据此可适当调整课程教学计划，设计相应教学知识点，补充现有内容涉及不足之处。

教学方式与教学团队方面，学习背景和兴趣偏好画像提示，67.7%的（21/31）学习者以同等学力方式入学，即这部分学习者在兼顾学习任务的同时，还需开展日常临床工作，将理论所学转化至临床应用。对此，应增加学习者临床实践和超声实操机会，使课程密切联系临床，根据临床需求跟踪前沿技术和研究热点。实践中，可与超声医学相关教学团队联合带教，安排学习者到超声科跟随专家出诊，熟悉诊断要点；在超声介入手术室观摩，对穿刺过程进行讲解、指导和答疑[23]。尽管3类不同兴趣偏好的学习者对教学内容的关注度不同，但皆认为教师教学态度和课程授课方式影响其学习效果，提示本课程教学团队要注重行之有效的授课方式、注重教师自身专业发展。此外，两类学习者反映出更多的学习效果影响因素，包括教师语言风格、学历学识和课程考勤制度、互动交流，但这两类学习者对课程内容的关注度相对较低。所以，需要教学团队针对不同兴趣偏好类型学习者开展精准化教学设计，注重学习者非认知能力的发展，促进学习动机和学习投入，进而激发学习兴趣、提升教学效果[24]。

学习效果预测方面，开展有效的教育质量监测可以为教学实践提供进一步数据支持和参考[25]。研究构建了学习者课程成绩的分类预测模型，虽有待扩展数据集进一步仿真验证，但研究实现了学习效果预测模型的概念验证，为今后在教学管理过程中关注学习行为、开展学情诊断和成绩预警奠定了基础。

4.2　研究局限性

可获取学习者数据的局限性。因研究执行年度选课人数较少，故可获取的样本量较小，且合成数据会引入部分噪声，所选数据分析方法和机器学习算法表现相对受限。在条件允许的情况下，可开展更多临床医学专业课程的实证研究，或针对某专

业或年级的学生进行画像建模。此外，当前课程虽以在线授课方式展开，但无法获取大规模在线学习平台上学习者浏览与访问课程资源、参与互动情况、预习与测试等能够表征学习行为和效果的更多数据。且因课程暂时取消了实习部分，无法记录学习者超声演示表现情况，导致部分画像标签缺失。

学习者画像的动态实时性不足。随着人工智能技术与教育场景的深度融合，用户画像的表现和应用更趋向于个性化与动态实时性。个性化要求模型具备优质的表征，动态实时性要求模型满足期望频率的更新需求[26]。当前构建的画像模型虽已覆盖完整的教学周期，但因各数据来源系统相对独立，不利于向实践应用转化。在后续研究中，仍需聚焦原始数据样本进行数据模型表征与高频率抽取更新，在已有系统中设计画像相关模块，促进画像模型由概念验证走向临床实际教育教学应用。

利益冲突　所有作者声明无利益冲突

作者贡献声明　刘蕾：论文构思、编写代码、撰写论文；王蕾、王晨：数据收集、预处理及统计分析；王云、高清：项目及教学内容指导、审订论文

参考文献

[1] 孙立会，刘思远，李芒. 面向2035的中国教育信息化发展图景：基于《中国教育现代化2035》的描绘[J]. 中国电化教育，2019(8)：1-8，43. DOI: 10.3969/j.issn.1006-9860.2019.08.001.

[2] 于方，刘延申. 大数据画像：实现高等教育“依数治理”的有效路径[J]. 江苏高教，2019(3)：50-57. DOI: 10.13236/j.cnki.jshe.2019.03.009.

[3] 刘毅，孙立，傅强，等. 模块化教学在超声引导神经阻滞教学中的应用探讨[J]. 中国继续医学教育，2020，12(23)：59-61. DOI: 10.3969/j.issn.1674-9308.2020.23.022.

[4] 马冬梅，徐桂萍. 超声引导神经阻滞在住院医师规范化培训中的应用[J]. 中华医学教育探索杂志，2016，15(6)：589-591. DOI: 10.3760/cma.j.issn.2095-1485.2016.06.011.

[5] 刘海鸥，孙晶晶，苏妍嫄，等. 国内外用户画像研究综述[J]. 情报理论与实践，2018，41(11)：155-160. DOI: 10.16353/j.cnki.1000-7490.2018.11.028.

[6] Antonenko PD, Toy S, Niederhauser DS. Using cluster analysis for data mining in educational technology research [J]. Education Tech Research Dev, 2012, 60(3): 383-398. DOI: 10.1007/s11423-012-9235-8.

[7] 翟鸣宇，程建，王苏桐，等. 基于K-prototype聚类的学生教育画像分析[J]. 大连理工大学学报(社会科学版)，2021，42(6)：22-31. DOI: 10.19525/j.issn1008-407x.2021.06.003.

[8] 许智宏，李彤彤，董永峰，等. 基于改进K-means算法的学生用户画像构建研究[J]. 河北工业大学学报，2022，51(3)：19-24. DOI: 10.14081/j.cnki.hgdxb.2022.03.003.

[9] Tian F, Wang S, Zheng C, et al. Research on E-learner personality grouping based on fuzzy clustering analysis [C]//2008 12th International Conference on Computer Supported Cooperative Work in Design. April 16-18, 2008, Xi'an, China. IEEE, 2008: 1035-1040. DOI: 10.1109/CSCWD.2008.4537122.

[10] 孙发勤，董维春. 基于学习分析的在线学习用户画像研究[J]. 现代教育技术，2020，30(4)：5-11. DOI: 10.3969/j.issn.1009-8097.2020.04.001.

[11] 唐佳诚，文欣秀，王家辉，等. 用户画像分层教学模型研究[J]. 实验室研究与探索，2018，37(12)：196-200. DOI: 10.3969/j.issn.1006-7167.2018.12.045.

[12] 胡小勇，林梓柔. 精准教研视域下的教师画像研究[J]. 电化教育研究，2019，40(7)：84-91. DOI: 10.13811/j.cnki.eer.2019.07.011.

[13] 程小恩，温川飙，程爱景，等. 基于网络课程的教师教学画像数据模型构建[J]. 成都中医药大学学报(教育科学版)，2018，20(4)：94-96.

[14] 黄文林. 基于学生画像分析的高校精准思政探索[J]. 东北大学学报(社会科学版)，2021，23(3)：104-111. DOI: 10.15936/j.cnki.1008-3758.2021.03.014.

[15] 师亚飞，彭红超，童名文. 基于学习画像的精准个性化学习路径生成性推荐策略研究[J]. 中国电化教育，2019(5)：84-91. DOI: 10.3969/j.issn.1006-9860.2019.05.013.

[16] 黄泰华，张涛，王磊. 多源数据融合视角下的大学生“消费-学业-社交”画像构建研究[J]. 农业图书情报学报，2022，34(7)：76-87. DOI: 10.13998/j.cnki.issn1002-1248.22-0165.

[17] 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. DOI: 10.3969/j.issn.1006-9860.2020.12.006.

[18] 肖君，乔惠，李雪娇. 基于xAPI的在线学习者画像的 构建与实证研究[J]. 中国电化教育，2019(1)：123-129. DOI: 10.3969/j.issn.1006-9860.2019.01.018.

[19] 陈良臣，傅德印. 面向小样本数据的机器学习方法研究综述[J]. 计算机工程，2022，48(11)：1-13. DOI: 10.19678/j.issn.1000-3428.0065347.

[20] 陈香萍，张奕，庄一渝，等. PROBAST：诊断或预后多因素预测模型研究偏倚风险的评估工具[J]. 中国循证医学杂志，2020，20(6)：737-744. DOI: 10.7507/1672-2531.201910087.

[21] 李雪娇. 基于学习者画像的在线学习支持服务策略设计研究[D]. 上海：华东师范大学，2019.

[22] 徐琦. 在线教育环境下基于学习者画像的学习成绩预测实践研究[D]. 黄石：湖北师范大学，2022.

[23] 贾东林，李水清，蒋洁，等. 多元化联合教学模式在超声引导下疼痛介入治疗中的应用[J]. 基础医学与临床，2022，42(1)：197-200. DOI: 10.16352/j.issn.1001-6325.2022.01.032.

[24] 白雪梅，顾小清，尹欢欢，等. 数据驱动精准教学：实践路径、感知理解与现实困境[J]. 电化教育研究，2022，43(4)：77-84. DOI: 10.13811/j.cnki.eer.2022.04.011.

[25] 张驰，李彦青，刘德平，等. 临床医学研究生学习行为的预测模型研究：线性回归和机器学习的对比分析[J]. 中华医学教育探索杂志，2021，20(3)：350-355. DOI: 10.3760/cma.j.cn116021-20200228-00439.

[26] 葛迪，吴彦文，刘三女牙. 低成本自进化的学习者画像模型研究[J]. 计算机工程与应用，2023，59(11)：141-150. DOI: 10.3778/j.issn.1002-8331.2202-0240.

（收稿日期：2023-02-05）
（本文编辑：张学颖）