

# 基于在线学习行为数据的学习者画像技术研究

李成

四川现代职业学院, 四川 成都 610207

**摘要：**随着 AI 与大数据技术的深入应用, 职业教育领域数字化转型加速演进。学习者画像技术作为个性化教育发展的关键, 本文通过分析在线学习行为数据, 洞察学生学习习惯、兴趣及能力, 助力精准教学和学习路径定制。有效解决了在实际应用中面临数据采集难、算法偏差大、隐私保护及平台集成难等挑战。基于在线学习行为数据的画像技术遵循了数据采集的全面性、准确性、实时性和隐私保护原则, 采用聚类、分类、关联规则挖掘等算法, 确保画像精准。该技术的研究成果将应用于个性化教学、学习评估等领域, 为智慧教育生态系统构建奠定基础。

**关键词：**在线学习; 行为数据; 学习者画像

## Research on learner portrait technology based on online learning behavior data

Li Cheng

Sichuan Modern Vocational College, Chengdu, Sichuan 610207

**Abstract：** With the in-depth application of AI and big data technologies, the digital transformation of the vocational education field is accelerating. As a key factor in the development of personalized education, learner profiling technology analyzes online learning behavior data to gain insights into students' learning habits, interests, and abilities, thereby facilitating precise teaching and customized learning paths. It effectively addresses challenges faced in practical applications, such as difficulties in data collection, significant algorithm biases, privacy protection concerns, and platform integration difficulties. The profiling technology based on online learning behavior data adheres to the principles of comprehensiveness, accuracy, real-time nature, and privacy protection in data collection. It employs algorithms such as clustering, classification, and association rule mining to ensure the accuracy of the profiles. The research findings of this technology will be applied in areas such as personalized teaching and learning assessment, laying a foundation for the construction of a smart education ecosystem.

**Keywords：** online learning; behavioral data; learner profiling

## 引言

随着生成式人工智能<sup>[1]</sup>与大数据技术<sup>[2]</sup>的迅猛发展, 职业教育领域正经历着深刻的数字化转型。这一变革不仅为我国职业教育带来了前所未有的发展机遇, 也伴随着诸多挑战。在信息爆炸的时代, 信息过载<sup>[3]</sup>与知识迷航现象<sup>[4]</sup>日益严重, 学生容易在学习过程中迷失方向, 导致学习进程紊乱, 效率低下。因此, 如何确保每位学习者在虚拟环境中有效吸收知识, 保持持久的学习热情, 成为职业教育领域亟待解决的关键问题。

学习者画像作为个性化教育的核心驱动力, 其重要性日益凸显。通过大数据分析 with 人工智能技术, 学习者画像能够深入洞察每位学生的学习习惯、兴趣偏好及能力水平, 精准识别其学习风格, 有效预测学习表现。这不仅为教师提供了科学调整教学策略的依据, 更为学生量身定制了个性化的学习路径, 确保了教学内容的针对性和有效性。因此, 深入探索并构建基于在线学习行为数据的学习者画像, 对于提升在线教育质量、优化学习体验、推动我国职业教育向智能化、个性化方向发展具有重要意义。

## 一、研究现状

### (一) 理论基础

用户画像理论<sup>[5]</sup>起源于市场营销领域, 由阿兰·库珀提出, 后逐渐引入教育领域, 形成学习者画像的概念。学习者画像的构

建主要基于三大理论支柱: 成人学习理论<sup>[6]</sup>、用户画像理论以及教育心理学<sup>[7]</sup>相关理论。

1. 成人学习理论: 如诺尔斯的 Andragogy 理论和塔夫的我指导学习理论, 强调成人学习者需要自我安排和指导学习的能力, 以及以学习者为中心的教学理念。这些理论为学习者画像的

基金项目: 2023 年度四川省教育信息技术研究课题“基于大数据的个性化学习路径推荐技术研究”(项目编号: DSJZXKT324)。

构建提供了关于成人学习特点和需求的重要视角。

2.教育心理学理论：关注学习者的认知过程、学习动机和学习风格。认知过程理论揭示了学习者如何接收、处理和应用信息；学习动机理论探讨了激发和维持学习者学习动力的因素；学习风格理论则指出了不同学习者在学习方式上的偏好和差异。这些理论为学习者画像的构建提供了丰富的心理学依据，有助于更深入地了解学习者的内心世界，构建出更加精准的学习者画像。

### （二）应用实践

学习者画像技术已在个性化教学支持、学习评估与反馈等领域得到初步应用。一些先进的学习平台和课堂教学工具通过收集和分析学习者的学习行为、学习成效、学习习惯等数据，构建出每位学习者的个性化画像。这些画像不仅记录了学习者的基本信息和学习轨迹，还深入分析了其学习风格、学习动机<sup>[8]</sup>和心理特征<sup>[9]</sup>。

教师在课中和课后通过对学习者画像的分析，可以全面了解学习者的学习成效和进步情况，为学习评估提供客观、全面的数据支持。同时，根据画像反馈，教师可以及时调整教学策略和教学内容，以满足不同学习者的个性化需求。这种基于数据的精准教学模式有效提高了教学效率和质量，增强了学生的学习体验和满意度。

### （三）主要问题

尽管学习者画像技术在教育领域取得了一定成果，但仍面临诸多挑战。

- 1.数据采集艰难。数据来源广泛且格式多样，包括学习平台、社交媒体、移动设备等，数据质量和完整性难以保证。
- 2.算法偏差。不同算法适用于不同场景和数据类型，选择不当可能导致画像结果不准确或存在偏差。
- 3.数据隐私。涉及大量个人数据的收集和處理，如何确保数据隐私性和安全性成为重要伦理和法律问题。
- 4.平台集成复杂性。不同平台数据格式、结构和质量差异大，增加了数据整合与处理的难度。

## 二、学习者画像构建的原则

为确保基于在线学习行为数据的学习者画像的准确性和有效性，需遵循以下原则：

### （一）学习行为数据的全面性

学习者画像应全面反映学习者的各种特征，包括基本信息（如性别、年龄、年级、专业等）、学习行为数据（如学习时长、学习资源偏好、学习进度等）、学习风格<sup>[10]</sup>数据（如认知风格、学习动机、学习偏好等）以及学习成效数据（如考试成绩、作业完成情况等）。同时，还需关注学习者的主观感受和心理特征，如学习态度、情绪状态等，以构建出更加全面和立体的学习者画像。

### （二）学习行为数据的准确性

确保收集到的数据准确无误是构建精准学习者画像的基础。需采取多源数据验证、数据清洗<sup>[11]</sup>和预处理等手段，提高数据的

准确性。通过整合来自不同渠道的数据并进行交叉验证<sup>[12]</sup>，可以进一步提高数据的可靠性。同时，对数据进行清洗，去除无效值和缺失值，并采用合理的插值或删除方法处理缺失数据，以保证数据的完整性和可靠性。

### （三）学习行为数据的实时性

学习者的行为和特征在不断变化，因此数据的实时性对于构建精准学习者画像至关重要。需利用实时数据收集和分析技术动态更新画像。通过在线学习平台的行为日志、可穿戴设备<sup>[13]</sup>采集的生理数据等实时数据源，可以实时追踪学习者的学习进度、资源访问情况和互动讨论等行为，确保画像的时效性和准确性。

### （四）学习者个人的隐私保护

在构建学习者画像的过程中，隐私保护是核心原则。需严格遵循国家法律法规及行业准则，对收集到的数据进行严格的匿名化处理，确保学习者的个人信息不被泄露。同时，增强透明度，明确告知学习者画像构建的目的、使用范围及可能产生的影响，并在取得学习者明确同意的前提下进行数据使用。通过全面保护学习者的隐私权益，构建安全可信的学习环境。

## 三、画像构建的方法

### （一）整体框架

学习者画像的构建流程包含四个核心阶段：数据收集、数据处理、特征提取与模型验证。

- 1.数据收集。从学习管理系统、在线平台等多渠道获取学习者的全面数据，如登录时间、学习时长、课程完成度、作业提交及测试成绩等。
- 2.数据处理。包括数据清洗（去重、剔除无效及缺失值）、整合（跨平台数据统一）及标准化（消除量纲差异），确保数据质量。
- 3.特征提取。根据画像目的选取关键特征，构建标签体系<sup>[14]</sup>，如学习态度、时间管理、学习效率等，并细化至具体行为指标。
- 4.模型验证。通过数据集划分、性能评估指标应用及模型优化，确保画像模型准确可靠。

### （二）数据收集

全面覆盖学习者的在线行为、基本信息、心理特征及学习成效等多维度数据。利用问卷调查、平台接口、在线测试及可穿戴设备等技术手段，确保数据的丰富性和代表性。

### （三）数据处理

精细化处理是画像准确性的基础。清洗过程剔除杂质，整合技术统一格式，标准化处理则确保数据可比性。每一步都力求数据质量最优化，为后续分析打下坚实基础。

### （四）维度划分

学习者画像的维度细分为五大部分：基本信息包括了性别、年龄、年级等背景信息。学习行为包括了记录学习习惯、频率、进度及资源访问情况。学习风格包括了反映认知风格、动机及偏好，如视觉型或听觉型学习者。学习成效包括了评估学习成果，

如考试成绩、作业完成情况。心理特征包括了探索情绪状态、学习态度等深层次特征。这些维度共同构成学习者画像的丰富内容，为个性化教学提供全方位视角。

#### （五）算法选择

算法的选择直接关系到画像的精准度和效率。常用的算法包括：聚类算法，如 K-means，用于学习者群体划分。分类算法，如决策树<sup>[10]</sup>、逻辑回归，预测学习者特征或行为。关联规则挖掘可以发现学习行为间的内在联系。协同过滤可以用于推荐相似学习资源或伙伴。深度学习，如神经网络，自动学习复杂特征表示。不同算法适用于不同场景，综合应用以提升画像质量。

#### （六）模型验证

通过科学划分数据集、严格测试性能指标及针对性优化，确保模型准确性。数据集的合理划分、多样化的性能评估指标及模

型参数的精细调整，共同构成了模型验证的完整流程。这一过程体现了科学严谨的方法论，保障了学习者画像技术的有效性。

## 四、结语

基于在线学习行为数据的学习者画像技术，正推动职业教育向智能化、个性化方向发展。该技术不仅为个性化教学、智能推荐及精准评估提供了技术支撑，还在职业培训、终身学习等领域展现出巨大潜力。随着大数据与 AI 技术的不断进步，学习者画像的构建将更加智能、精细，为构建智慧教育生态系统贡献力量。这不仅是教育个性化的关键一步，更是提升职业教育质量、培养未来人才的重要驱动力。

## 参考文献

- [1] 武林娜, 宋恺, 王淞鹤. 生成式人工智能对个人信息安全的挑战及应对策略 [J]. 信息通信技术与政策, 2024, 50(01): 13-18.
- [2] 杨现民, 唐斯斯, 李冀红. 发展教育大数据: 内涵、价值和挑战 [J]. 现代远程教育研究, 2016, 50-61.
- [3] 杨永真. 信息过载研究综述 [J]. 财讯, 2019, 32-34.
- [4] 刘艳朵, 刘敏. 信息迷航研究脉络与启示 [J]. 内蒙古科技与经济, 2022, 87-90.
- [5] Pujahari A, Sisodia D S. Item feature refinement using matrix factorization and boosted learning based user profile generation for content-based recommender systems. Expert Systems with Applications, 2022, 56-58.
- [6] 方明建. 基于成人学习理论的教师信息化教学能力培养原则和模式研究 [J]. 现代教育技术, 2024, 33-36.
- [7] 蔡大生. 高等教育心理学. 湖南人民出版社, 2010.
- [8] 曾小军; 李双玲; 浅谈激发大学生学习动机的若干策略 [J]. 新疆教育学院学报; 2008年02期.
- [9] 姜琳, 钟志兵, 冯振军, 等. 心理素质内涵、结构及特征述评 [J]. 科技视界, 2016, (27): 237+263.
- [10] 杨文华. 中师不同专业学生认知风格差异与教学策略研究 [D]. 云南师范大学, 2004.
- [11] J. Van den Bossche, R. Bostoen, and F. Ongenae. A systematic review on data cleaning for fraud detection. Knowledge and Information Systems, 2019, 139-163.
- [12] 马克·德·罗伊 (Mark de Rooij) 和沃特·维达 (Wouter Weeda). 交叉验证: 每个心理学学者都应了解的方法. 心理科学方法与实践进展, 2020.
- [13] James J. Collins (哈佛大学) 和 Can Dincer (弗赖堡大学) 等人. End-to-end design of wearable sensors. Nature Reviews Materials, 2022
- [14] 姚华彦, 张鑫金, 何萍. 基于大数据的患者画像标签体系构建方法及应用研究 [J]. 中国卫生信息管理杂志, 2019, 16(06): 667-671.
- [15] 孙明檀. 基于多任务学习的软决策树方法 [D]. 云南财经大学, 2024.