

# 基于在线学习行为的学习者画像构建

郭煜, 杨艳青

北京工业大学信息化建设与管理中心, 北京

收稿日期: 2023年1月24日; 录用日期: 2023年2月21日; 发布日期: 2023年2月28日

## 摘要

学习者画像是大数据时代下的产物, 其通过对学习者相关数据的分析来呈现学习者的学习特征, 从而更好地为提升在线学习质量服务。如何挖掘分析学习者的相关数据来构建学习者画像是当前研究普遍关注的问题。本研究首先以日新学堂在线学习平台中的在线学习行为数据为切入点, 基于交互视角对已有的在线学习行为数据进行分类, 并据此划分学习者画像的维度。其次, 在进一步明确学习者画像的标签后, 通过采用K-means聚类方法形成三类学习者画像, 并进一步将其命名为: 高沉浸性学习者、中沉浸性学习者、低沉浸性学习者, 对其画像进行描述, 呈现三类学习者的在线学习行为特征。最后, 提出学习预警是学习者画像的重要应用方向。

## 关键词

交互, 在线学习行为, 学习者画像

# Construction of the Learner Portrait Based on Online Learning Behavior

Yu Guo, Yanqing Yang

Information Construction and Management Center, Beijing University of Technology, Beijing

Received: Jan. 24<sup>th</sup>, 2023; accepted: Feb. 21<sup>st</sup>, 2023; published: Feb. 28<sup>th</sup>, 2023

## Abstract

Learner portrait is a production in the era of big data. It presents the learning characteristics of learners through the analysis of learner-related data, so as to better serve to improve the quality of online learning. How to mine and analyze the relevant data of learners to construct learner portraits is a common concern of current research. This research takes the online learning behavior data in the Rixin Xuetang online learning platform as the starting point, and classifies the ex-

isting online learning behavior data based on the interactive perspective, and divides the dimensions of the learner portrait accordingly. After further clarifying the label of learner portraits, three types of learner portraits are formed by using k-means clustering method, and further named as high immersion learners, medium immersion learners and low immersion learners. Their portraits are described to present the learning behavior characteristics of the three types of learners. Finally, it is proposed that learning early warning is an important application direction of learner portrait.

## Keywords

Interaction, Online Learning Behavior, Learner Portrait

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着信息科学技术的发展, 信息化成为现代社会的重要标志, 高校也在积极推进信息技术与教学的融合, 推进在线学习模式在高校的展开, 使在线学习成为教学新常态。然而在线学习中还存在的诸多问题, 如学生学习倦怠[1]、参与性低、难以进入深度学习[2]、学习投入不足[3]等, 都将导致学生的高流失率和低完成率[4], 让在线学习质量面临着诸多挑战。

学习者画像作为学习分析技术的典型应用, 其潜力被逐渐挖掘, 价值被日益证实。学习者画像可以通过对教育大数据的分析与处理来挖掘隐性信息, 来为在线学习提供支持服务。对于学习者个体而言, 学习者画像能精准判断学习者的学习状态[5]; 能反映风险学习者的行为表现和学习路径特征[4]; 能促进学习者自身认知与发展[6]。对于教育管理者而言, 学习者画像能实现自适应精准推荐, 驱动教育管理者循证决策。

现有的学习者画像主要是从学习者数据入手来刻画学习者的特征, 常见的学习者数据类型有人口学数据、心理数据、行为数据和学习结果数据等。其中, 行为数据具有特殊的价值, 它隐含着学生行为特征, 对于分析学习过程、探求行为模式、分析其与学习效果的关系、预测学习结果意义深远[7]。因此, 本研究的研究重心便在于学生行为数据, 将以在线学习行为为切入点, 通过对学习者在线学习行为的跟踪和分析, 来刻画学习者在在线学习过程中所隐含的行为特征, 构建基于在线学习行为数据的学习者画像。具体而言, 本研究主要关注的问题是: 1) 学习者的在线学习行为类型有哪些? 2) 如何基于在线学习行为来构建学习者画像?

## 2. 在线学习行为的相关研究

在线学习行为是学习者与在线学习环境交互过程中产生的行为, 既关乎学习者的学习活动, 又关乎学习者的社会化特征。本研究的目标是构建基于在线学习行为的学习者画像, 从具体的在线学习行为中挖掘学习者的行为特征。为此, 本研究首先要厘清学习者的在线学习行为类型, 对在线学习行为进行分类。

### 2.1. 在线学习行为分类

在线学习行为是复杂多样的, 目前研究者们对于在线学习行为分类的方法主要是理论演绎, 通过对

理论的梳理来确定学习行为的层次和类别, 并进一步明确学习行为的特征和表征方式[8]。典型的在线学习行为分类如吴林静等根据学习者与网络学习空间的学习系统、学习资源、人的相互关系, 将学习者的在线学习行为分为四类: 独立学习行为、系统交互行为、资源交互行为和社会交互行为[9]; 王丽娜根据学习者的交互情况将学习者的在线学习行为分为两类: 个性化交互学习行为和社会性交互学习行为两类[10]; 彭文辉结合学习行为的 OCCP 层次分类模型, 建立在线学习行为的 S-F-T 三维分类模型, 从三个维度对学习者的在线学习行为进行分类: 行为结构层次维度、行为方式维度、行为功能维度[11]。

以上研究者在对在线学习行为进行分类时或从与教学要素的交互角度, 或从个体性和社会性的交互角度, 或者结合已有框架从多角度对在线学习行为的分类进行理论演绎。虽然维度不同, 但研究者间的观点仍然存在共性。例如以上研究者在谈及在线学习行为分类时, 均提到了“交互”的概念, 在此处“交互”概念或是狭义上的人与人的互动, 或是广义上的人与环境的互动。

## 2.2. 基于交互视角的在线学习行为分类

交互即相互作用, 在在线学习中, 交互是学习者正确建构学习内容, 而与学习过程中的相关要素相互交流和作用的过程[12]。在线学习行为的本质便是学习者与学习要素的交互, 是一种信息交流活动。本研究结合在线学习行为的要素, 借鉴吴林静等人的网络空间结构模型和对在线学习行为的分类和度量[9], 将在线学习行为分为: 系统交互行为、资源交互行为、人际交互行为(见表 1)。系统交互行为是学习者与在线学习平台系统之间交互产生的行为, 如参与签到、提交作业、提交测验等; 资源交互行为是学习者与在线学习平台上资源之间交互产生的行为, 如观看音视频等; 人际交互行为是学习者与在线学习平台的其他个体之间交互产生的行为, 如论坛发帖回帖、参与抢答、参与评分等。此处基于交互视角对在线学习行为进行分类的意义在于明确在线学习行为类型, 从而为学习者画像的研究奠定数据分析的基础。

**Table 1.** Classification of online learning behaviour

**表 1.** 在线学习行为分类

在线学习行为	典型行为
系统交互行为	参与签到、提交作业、提交测验
资源交互行为	观看音视频
人际交互行为	论坛发布回复评论、参与抢答、参与评分

## 3. 基于在线学习行为的学习者画像构建

学习者画像的价值在于使得隐性信息得到挖掘。在明确在线学习行为分类后, 更重要的是基于此进行学习者画像的构建, 从而更好地分析和利用学习者数据, 为实现学习者的风险干预和个性化学习、提高在线学习质量奠定基础。本研究将从明确画像目标、划分画像维度、确定画像标签、进行数据分析、验证画像准确性、形成群体画像这几个步骤来进行学习者画像的构建。

### 3.1. 明确画像目标

自 2020 年初新冠疫情爆发以来, 日新学堂在线学习平台开展广泛的在线学习活动, 在此过程中产生了大量的在线学习行为数据。学习者画像的目标便在于通过对在线学习行为数据的挖掘, 描绘学习者的行为特征, 从而便于教与学利益相关者采取针对性的措施, 为提高在线学习质量服务。

### 3.2. 划分画像维度

学习者画像是基于对学习行为数据的分析建立的。在线学习中, 学习者的数据一般而言可以分为人口

学数据、心理数据、行为数据、学习结果数据等多种类型, 已有研究对学习画像维度的划分也主要一般主要是从这几个维度展开。

本研究中学习画像生成于在线学习行为, 且基于在线学习行为进行构建, 因此画像维度也应基于在线学习行为的分类维度中产生。前文中基于交互视角对在线学习行为进行分类, 画像的维度也将从交互视角进行划分。具体而言, 本研究将从系统交互、资源交互、人际交互这三个维度来划分学习画像维度(图 1)。其中“系统交互维度”旨在探究学习者与在线学习平台系统间交互的特征; “资源交互维度”旨在探究学习者与在线学习平台资源间交互的特征; “人际交互维度”旨在探究学习者与学习中的其他个体间交互的特征。

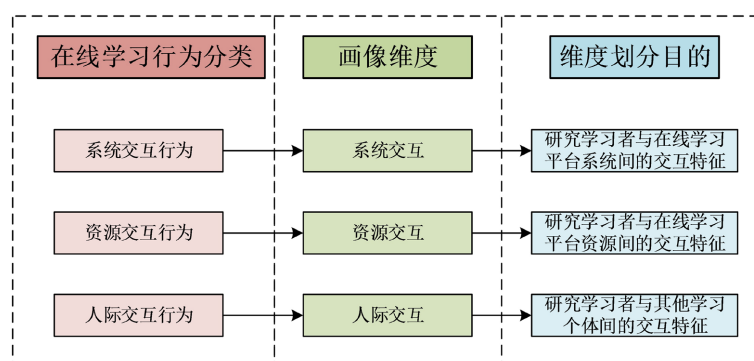


Figure 1. Portrait dimension  
图 1. 画像维度

### 3.3. 确定画像标签

标签是一种简要标示, 学习画像中标签的作用在于用简洁、精准的文本来标示学习者的特征。学习画像标签从数据中产生, 画像标签的设计要结合数据的实际情况进行。

数据挖掘的特性在于针对有意义的数据进行深入探究。本研究在进行画像标签设计时重点考虑能反映学习者学习频次、学习时长、学习时间点的特征。结合日新学堂在线学习平台的已有数据和画像的维度, 本研究划分的画像标签及数据指标如下表 2 所示。具体而言, 在系统交互维度将画像标签设置为出勤频次和时间节点, 可以了解学习者的学习频次情况和学习时间点情况; 在资源交互维度将画像标签设置为学习时长和复习程度, 可以了解学习者的学习时长情况和复习情况; 在人际交互维度将画像标签设置为互动频次和互动深度, 可以了解学习者的互动情况。这些标签综合反映了学习者的学习特征。

Table 2. Portrait labels and data indicators

表 2. 画像标签及数据指标

画像维度	画像标签	数据指标
系统交互	出勤频次	参与签到次数、提交测验次数、提交作业次数
	时间节点	参与签到时间点、提交测验时间点、提交作业时间点
资源交互	学习时长	音视频观看时长
	复习程度	音视频观看反白比
人际交互	互动频次	总讨论数、发表评论次数、回复讨论次数、参与抢答次数、参与评分次数、参与问卷次数
	互动深度	发表评论次数/总讨论数、回复讨论次数/总讨论数、获赞数

### 3.4. 进行数据分析

在明确画像的维度和标签后, 还要结合实际数据进行分析, 构建具体的学习者画像。本研究将基于日新学堂在线学习平台中的在线学习行为数据来展开研究。在学习行为数据的选择上, 主要选择日新学堂在线学习平台上的课程《毛泽东思想与中国特色社会主义理论体系概论》中的在线学习行为数据。该课程是日新学堂在 2020 年春季学期所开设的公共必修课程, 其课程资源较为多样, 所积累的在线学习数据较为丰富。已有数据信息如表 3 所示。

**Table 3.** Data and information of the selected courses

**表 3.** 所选课程数据信息

课程名称	《毛泽东思想与中国特色社会主义理论体系概论》
班级设置	三个班级, 共 337 名学生
课程资源	12 个课程章节, 章节任务点 30 个, 视频 6 个, 音频 88 个, 文档 40 个, 其他资料 56 个
学习行为	参与签到、提交测验、提交作业、观看音视频、发表讨论、回复讨论、参与抢答、参与评分、参与问卷

在明确数据范围后, 本研究进一步对数据进行预处理。数据预处理是数据分析的前期准备工作, 是进一步研究开展的基础。在数据处理工具上, 本研究选择 Excel、SPSS 作为数据处理的主要工具。在数据处理方式上, 本研究通过清洗来筛除无关数据和极值数据, 通过归一化来统一数据形式。经过预处理后, 学习者的不同学习行为被量化成画像标签的具体数值, 数据变得统一完整, 便于下一步的分析。

学习者画像的本质便是对不同类型学习者的行为特征的描述, 因此确定学习者的类型对本研究而言至关重要。目前, 基于学习行为类型进行学习群体分类日益成为在线学习者研究的重要内容。相关研究者在探究基于学习者学习行为分类时主要采用 K-means 聚类方法, 将学习者进行聚类。为了了解不同群体学习者的学习行为特征, 本研究在借鉴相关研究的基础上, 也将利用 SPSS 软件, 采用 K-means 聚类方法对学习群体进行聚类。具体的分析流程如下:

首先, 本研究通过使用手肘法, 发现在  $k = 3$  时出现明显拐点, 即确定最优  $k$  值为 3, 当聚类数为 3 时聚类效果最好。此时共形成三个类簇, 即形成三个学习群体。

然后, 本研究通过对三类学习群体的画像标签值进行均值分析(见表 4), 以区分不同类型的学习者。可以发现学习者群体 2 的画像标签值均为最高, 学习者群体 1 的画像标签值均为最低, 学习者群体 3 的画像标签值介于二者中间。

**Table 4.** Mean values of portrait label values

**表 4.** 画像标签值均值

群体	人数	出勤频次	时间节点	学习时长	复习程度	互动频次	互动深度
1	40	1.29	1.29	0.15	0.08	0.96	0.04
2	141	2.61	2.61	0.39	0.13	2.91	0.32
3	156	2.14	2.14	0.29	0.09	1.47	0.11

最后, 本研究根据均值分析的结果来对学习群体进行命名。本研究在参考已有研究者对学习者的分类上, 借鉴王改花等人对学习者的聚类[13], 对本研究聚类的三类学习者进行命名和特征描述。具体而言, 本研究将学习者群体命名为高沉浸性学习者、中沉浸性学习者、低沉浸性学习者。高沉浸性学习者

的在线学习行为表现最好, 低沉浸性学习者的在线学习行为表现最差, 中沉浸性学习者的在线学习行为表现处于二者之间。

### 3.5. 验证画像准确性

画像的准确性是指学习者画像具有明确的区分度, 能反映不同类型学习者的特征。学习者画像是否准确关系着相应举措能否精准施行, 关系着在线学习质量能否有效提升。在明确学习者画像的维度和标签后, 本研究应进一步验证画像的准确性。

具体而言, 本研究主要通过验证学习者画像标签和学习成绩的相关性来说明画像的准确性。本研究通过在 SPSS 中对画像标签和学习成绩进行正态性检验, 其显著性明显, 符合正态分布(见表 5), 因此可以进行 Pearson 相关性分析。Pearson 相关性分析结果显示, 六项画像标签均与学习成绩呈显著性相关关系(见表 6), 即在线学习行为和学习成绩呈显著相关关系。这说明本研究所设置的画像标签是准确的, 能反映出不同成绩水平的学习者的在线学习行为特征。

**Table 5.** Test of normality

**表 5.** 正态性检验

	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
	统计	df	显著性	统计	df	显著性
出勤频次	0.078	337	0.000	0.968	337	0.000
时间节点	0.183	337	0.000	0.824	337	0.000
学习时长	0.07	337	0.000	0.953	337	0.000
复习程度	0.123	337	0.000	0.74	337	0.000
互动频次	0.065	337	0.001	0.978	337	0.000
互动深度	0.258	337	0.000	0.722	337	0.000
学习成绩	0.08	337	0.000	0.894	337	0.000

**Table 6.** Correlation analysis

**表 6.** 相关性分析

		出勤频次	时间节点	学习时长	复习程度	互动频次	互动深度	学习成绩
学习成绩	Pearson相关性	0.494**	0.215**	0.358**	0.163**	0.327**	0.366**	1
	显著性(双尾)	0.000	0.000	0.000	0.003	0.000	0.000	
	N	337	337	337	337	337	337	337

\*\*在置信度(双测)为 0.01 时, 相关性是显著的。

### 3.6. 形成群体画像

在形成三类学习者群体之后, 为更直观地呈现学习者的行为特征, 还应通过可视化的方式来对学习者的画像进行输出。本研究主要形成三类学习者群体画像。

高沉浸性学习者画像(见图 2、图 3): 该类学习者的人数少于中沉浸性学习者, 学习者的 6 类画像标签值均明显高于整体均值。在系统交互上, 该类学习者的出勤频次高, 学习任务完成率高; 时间节点早, 能尽早提交学习任务。在资源交互上, 该类学习者中的大部分人学习时长都属于长或一般水平; 大部分

学习者的复习程度都属于高或一般水平。在人际交互上, 该类学习者的互动频次较高, 互动深度也处于高或一般水平。

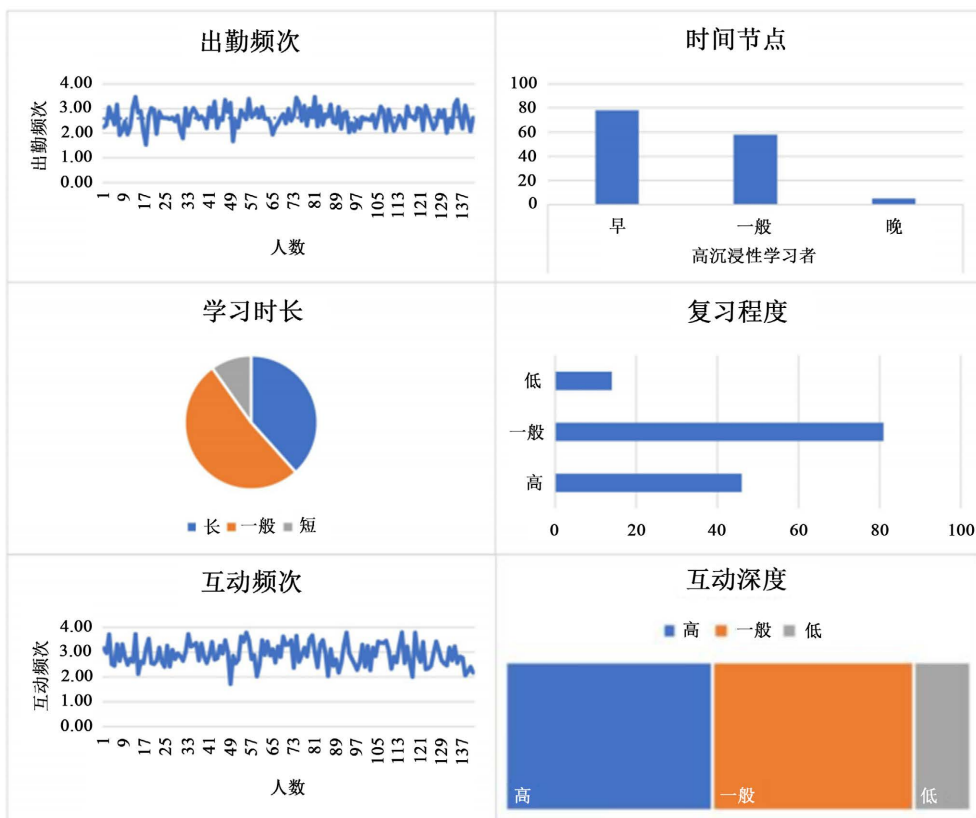


Figure 2. Label values of the highly immersive learners' portrait  
图 2. 高沉浸性学习者画像标签值

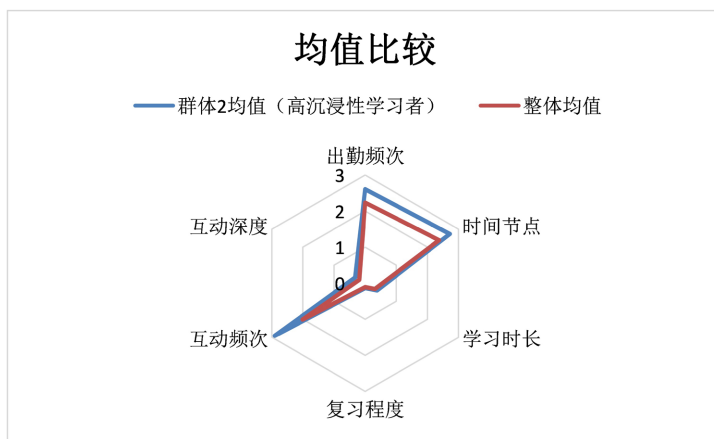


Figure 3. Comparison of the mean labels of highly immersive learners' portrait  
图 3. 高沉浸性学习者画像标签均值比较

中沉浸性学习者画像(见图 4、图 5): 该类学习者人数最多, 学习者的 6 项画像标签值与平均水平大致相同。在系统交互上, 该类学习者的出勤频次处于中间水平, 能完成大部分学习任务; 提交学习任务

的时间点比较适中, 没有尽早提交, 但能在规定时间内完成。在资源交互上, 该类学习者中的大部分学习者的习时长处于一般水平; 该类学习者的复习程度处于一般或较低水平。在人际交互上, 该类学习者的互动频次适中, 群体间的互动频次相差不大; 大部分学习者的互动深度处于较低水平。



Figure 4. Label values of the medium immersive learners' portrait

图 4. 中沉浸式学习者画像标签值

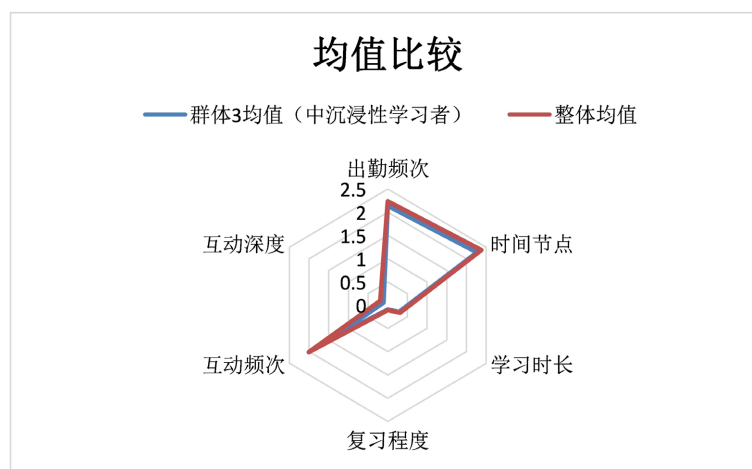


Figure 5. Comparison of the mean labels of medium immersive learners' portrait

图 5. 中沉浸式学习者画像标签均值比较



低沉浸性学习者画像(见图 6、图 7): 该类学习者人数最少, 学习者的 6 项画像标签值都显著低于平均水平。在系统交互上, 出勤频次低, 学习任务的完成度低; 该类学习者提交学习任务的时间点较晚, 容易出现拖延情况。在资源交互上, 该类学习者大部分学习者的学习时长偏短; 大部分学习者的复习程度处于一般或较低水平, 较少进行资源的复习。在人际交互上, 该类学习者的互动频次偏低, 较少参与人际互动; 具大部分学习者的互动深度都很低。

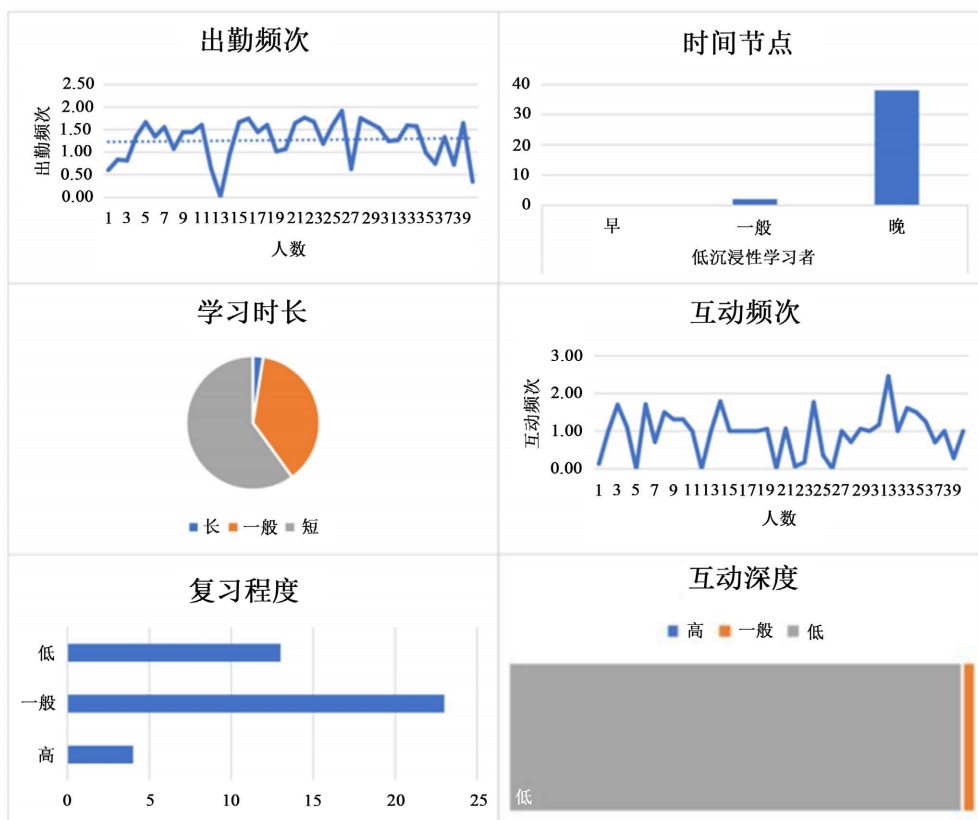


Figure 6. Label values of the lowly immersive learners' portrait  
图 6. 低沉浸性学习者画像标签值

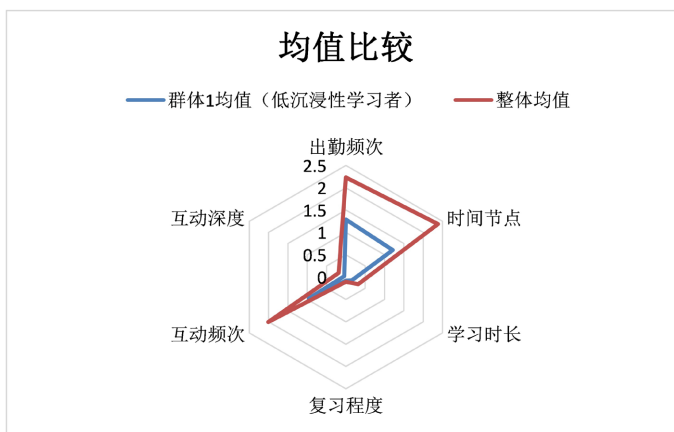


Figure 7. Comparison of the mean labels of lowly immersive learners' portrait  
图 7. 低沉浸性学习者画像标签均值比较

本研究通过以上对三个画像维度和六个画像标签的分析, 形成了三类学习者画像, 分别是高沉浸性学习者、中沉浸性学习者、低沉浸性学习者。这三类学习者具有明显的差别, 在前文中已经进行阐述。除了差别之外, 这三类学习者也存在一定的共性。最明显的便是在互动频次和互动深度上, 虽然这三类学习者有数量上的区别, 但是就总量而言, 三种学习者类型的总量都偏少, 很多学习者很少甚至是不参与人际互动。人际互动总量较低是该平台的学习者在线学习时所面临的突出问题, 这或许可以给我们提示, 应在增强平台的互动性, 如通过采取弹幕、参与交互获得积分等方式来鼓励学习者积极参与互动。

#### 4. 学习者画像的应用展望

本研究首先基于交互视角对学习者的在线学习行为进行分类, 并在此基础上探究学习者画像的维度划分。其次, 在明确学习者画像的维度之后, 结合日新学堂在线学习平台中的在线学习行为数据进行数据分析, 通过 K-means 聚类方法来形成三类学习者群体。最后, 通过对三类学习者群体的描述和可视化呈现, 形成三类学习者画像。

在形成学习者画像之后, 还应进行更深入的探究。学习预警是学习者画像的重要应用方向。预警要对事物进行前兆性预测, 学习预警的作用便在于对学习者的学习过程进行风险识别, 并通过相应的干预措施, 将学习风险降到最低, 从而提高学习成绩, 提升学习质量。在学习者画像的应用上, 可以基于学习者画像进行学习预警的研究, 根据不同类型学习者群体提出相应的预警措施和方案, 从而帮助学习者进行风险识别和过程干预, 辅助其提升在线学习效果, 提升在线学习质量。

#### 参考文献

- [1] 万力勇, 舒艾. 互动失衡对大学生在线学习倦怠的影响研究[J]. 现代远程教育, 2022(2): 34-41.
- [2] 舒莹, 姜强, 赵蔚. 在线学习危机精准预警及干预:模型与实证研究[J]. 中国远程教育, 2019(8): 27-34+58+93.
- [3] 杨秋燕. 大学网络通识课中学习者在线学习行为投入研究[D]: [硕士学位论文]. 重庆: 西南大学, 2019: 37.
- [4] 肖君, 乔惠, 李雪娇. 大数据环境下在线学习者画像的构建[J]. 开放教育研究, 2019, 25(4): 111-120.
- [5] 秦婷, 郑勤华. 联通主义学习社区个体知识贡献影响因素研究[J]. 现代远程教育, 2020(5): 52-61.
- [6] 莫尉. 学习者画像建构及应用研究[J]. 湖南理工学院学报(自然科学版), 2021, 34(3): 64-69.
- [7] 石芳华. 本科教育质量评价改革新视角: 学习投入度[J]. 现代教育管理, 2010, 242(5): 51-54.
- [8] 马志强, 苏珊. 学习分析视域下的学习者模型研究脉络与进展[J]. 现代远程教育, 2016, 166(4): 44-50.
- [9] 吴林静, 劳传媛, 刘清堂, 程云, 毛刚. 网络学习空间中的在线学习行为分析模型及应用研究[J]. 现代教育技术, 2018, 28(6): 46-53.
- [10] 王丽娜. 网络学习行为分析及评价[D]: [硕士学位论文]. 西安: 陕西师范大学, 2009: 19.
- [11] 彭文辉. 网络学习行为分析及建模[D]: [博士学位论文]. 武汉: 华中师范大学, 2012: 70-73.
- [12] 陈丽. 术语“教学交互”的本质及其相关概念的辨析[J]. 中国远程教育, 2004(3): 12-16+78-79.
- [13] 王改花, 傅钢善. 数据挖掘视角下网络学习者行为特征聚类分析[J]. 现代远程教育研究, 2018(4): 106-112.