

# 基于大数据技术的大学生学习画像构建

陈 会, 余 馨, 李琳琳, 吴苏徽, 蒋秀莲

(徐州工程学院, 江苏 徐州 221008)

✉1746599872@qq.com; 776434727@qq.com; 1317315180@qq.com;  
1109850278@qq.com; 447502622@qq.com



**摘 要:** 在信息社会, 各行各业的管理控制转变为以数据、信息为中心。在高等教育领域, 高校重视学生信息数据库的建设, 通过学生浏览信息的关键词、种类分布和浏览主题等多个维度构建学生画像向量空间模型。本文使用大数据技术构建学生学习画像基础模型框架, 研究学生学习画像在个性化学习、问题预警及辅助学校决策等方面的应用, 为高校提升学生培养质量提供参考。

**关键词:** 学习画像; 用户标签; 数据挖掘

**中图分类号:** TP311.13 **文献标识码:** A

## Construction of College Students' Learning Portrait based on Big Data Technology

CHEN Hui, YU Xin, LI Linlin, WU Suhui, JIANG Xiulian

(Xuzhou University of Technology, Xuzhou 221008, China)

✉1746599872@qq.com; 776434727@qq.com; 1317315180@qq.com;  
1109850278@qq.com; 447502622@qq.com

**Abstract:** In information society, it is data and information that manage and control all walks of life. In the field of higher education, universities attach importance to the construction of student information databases. The vector space model of student portraits is constructed through multiple dimensions such as keywords, type distribution and browsing topics of students' browsing information. This article uses big data technology to build a basic model framework for student learning portraits, and studies the aspects of student learning portraits in personalized learning, early warning of problems, and assistance in school decision-making, so as to provide references for colleges and universities to improve the quality of student training.

**Keywords:** learning portrait; user tags; data mining

### 1 引言(Introduction)

我国普通高等学校素质教育明确提出, 高校的教学任务在于不断提升学生的综合素质。信息社会下的大学生呈现个性化发展的趋势<sup>[1]</sup>, 他们的学习行为、特长偏好等也相对多样化。学校对学生的教育方式要适应学生的个性化发展需求, 以利于提升学生的综合素质, 为经济社会培养高质量人才。

当前国内在企业精准营销以及数据产品个性化推荐领域中, 对用户进行画像构建的较多。高校对学生的数据搜集、处理以及画像构建等尚不全面, 大多数画像构建通常停留在数据的描述可视化上<sup>[2]</sup>, 并未对学生的教育与改善学习效果起到明显作用。基于大数据技术的学生学习画像构建针对学生

不同个性发展的独立性及多样性, 重视学生在思维和行为上的差距, 突破对学生综合评价仅考虑学习成绩的局限性, 能更加全面地对学生进行评价及打分, 可以更好地引导学生, 挖掘学生潜能, 促进学生全面发展。本文探讨研究基于大数据技术的学生学习画像基础模型框架的构建, 以期在学生个性化学习、学生问题预警及辅助学校有关政策、决策的制定等方面提供数据驱动。

### 2 大学生学习画像(University students' learning portrait)

大学生学习画像是高校大学生在学习方面的虚拟代表, 是建立在一系列真实数据之上的目标用户模型。通过学生学

习数据收集分析了解学生,根据他们的目标、行为和属性的差异,将他们区分为不同的类型,然后从每种类型学生中抽取基本信息、内容偏好、学习风格和社交互动行为描述,就形成了一个人物原型即一个学生学习画像。根据数据的记录和描述性统计分析可得:在已知学生性别、年龄和专业的前提下,依据学生检索信息的内容、页面浏览的次数以及下载量,甚至包括在社交学习平台上资源转发频率和互动评论内容等,可以计算出每位在校大学生的学习状态,从而构建学生学习画像,预测学生学习成效,进而帮助教师更好地关注学生的学习状态和身心健康。此外,根据统计的数据记录能够辅助学校政策的制定,使得制定的政策更加人性化和专业化。

### 3 基于大数据技术的学习画像构建(Construction of learning portrait based on big data technology)

现行的用户画像主要运用网络流算法检验学生的学习状态,重点运用多层次聚类分析算法进行数据挖掘,运用多元回归分析和神经网络算法预测学生学习成绩及挂科率。鉴于一些高校对学生考评测评方式仅限于结构化数据的成绩分析,且存在数据挖掘意识不强等问题<sup>[3]</sup>,本项目对高校学生学习、消费、网络使用及生活等行为方面的结构和非结构化数据进行预处理和挖掘,构建学生学习画像,从而为学生个性化学习、学生问题预警、辅助学校决策等提供数据驱动,以加强高校优良学风建设。学生学习画像构建步骤如下:

第一步:将目标用户画像问题转化为学生学习画像问题。

学生学习画像分析本质上是从学生的角度思考问题,涉及若干学生用户群体、若干学生用户行为。网络课程通常有三种学习用户——存量学习用户、流失学习用户、潜在学习用户,涉及学生基本信息、学习目的、学习方式、学习态度、学习成效、学习评价和体验等,因此分门别类解释逻辑尤为重要。

第二步:宏观假设验证。

转化完问题后,需在拆解以前聚焦假设,先在宏观上对假设进行检验,有效避免无限拆解的错误。进行大方向检验,可以有效缩小怀疑范围。怀疑范围越小,后续对学生用户分析越精确<sup>[4]</sup>。当数据不足的时候,能改善数据质量,做出准确分析。

第三步:构建分析逻辑。

宏观验证以后,可基于已验证的结论,构建更细致的分析逻辑。在这个阶段,实际上已经把原本庞大的问题聚焦为一个一个小问题。学生学习画像构建框架可划分为三个层次:数据源层、数据挖掘层和业务应用层。数据源层需要对结构化和非结构化的数据进行提取;数据挖掘层则需对所提取的数据建模,针对所建立的模型和运算结果进行充分应用,是业务应用层的基础。学生学习画像构建框架具体如图1所示。

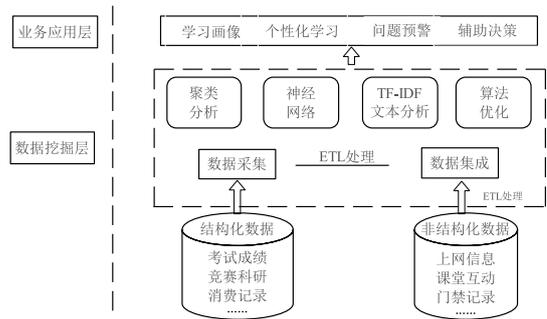


图1 学生学习画像构建框架示意图

Fig.1 The frame diagram of students' learning portrait

#### 3.1 学生学习画像的数据预处理

鉴于本文所需爬取的数据均存在于高校的学生信息数据库和各大学习网络平台上,且各大学习网络都提供了API,在数据爬取前申请key,以json形式返回文档,方便解析。通过各种学习、社交平台 and 上网流量监控,对学生的进行学习数据进行爬取。若数据呈结构化状态则直接提取,若数据呈非结构化状态则先对其进行赋值,再做数据无量纲化处理。利用模糊c均值聚类法和词云图过滤掉大量的文本信息及异常值,数据爬取时尽可能获取全量的学生学习数据,为教师对学生学习成绩的分析提供坚实的数据基础,如学生成绩数据、学生上网数据、学生消费数据、学生课堂行为数据及教师反馈数据等相关数据。数据来源如图2所示。提取相应的数据,量化后建立标签。

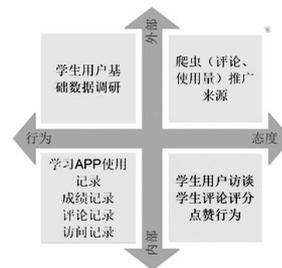


图2 数据来源

Fig.2 Data source

##### (1)基本信息标签

基本信息是指一个学生的基本信息和变更频率较低的代表性指标,此处提取学生的学号、姓名、性别、专业、班级及所关注的方向等,这些指标可以直接获取。

##### (2)学习风格标签

学习风格是学生用户非常重要的一个方面,学生对学习方式的偏好及喜爱程度是学生学习画像最重要的信息之一,是对用户和学习方式之间的关系进行深度刻画的重要标签,其中最典型的是视觉(影视网课)、听觉(语音录播)、言语(交流讨论)。

##### (3)内容偏好标签

内容偏好记录的是大学生学习、浏览、关注的内容。学生的浏览内容行为包括信息检索、页面浏览和资源下载等。由于这些浏览内容行为种类繁多且和不同的学习平台、不同的模块交互,不同时间进行不同操作,导致行为属性更加复杂。针对如何能够全面梳理,怎样才能集成抽取学生的内容偏好,可以按照图2所示的分类方法进行。

##### (4)社交互动标签

学生学习时会进行社交、分享等一系列互动活动,主要

有资源分享转发、收藏、互动评论等。在该过程中，有些学生会浏览比较陌生的领域知识，而有些内容要通过一定知识量和案例的引导才会促使学生更深入地学习。通过建立社交活动标签，可对不同专业的学生推送合理的学习资源，保证资源被学生最大化利用，使得投资回报率最大。该标签下多种不同属性的敏感度代表大学生对学习平台的敏感程度，也是典型的挖掘类标签。

学生标签分类示意图如图3所示。

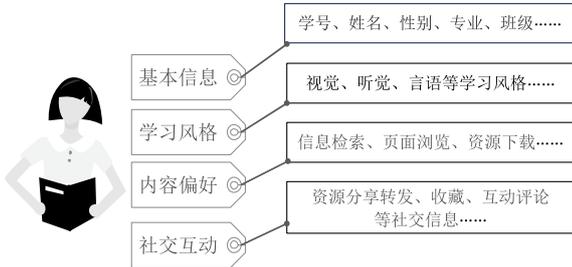


图3 学生标签分类示意图

Fig.3 Schematic diagram of student label classification

### 3.2 学生个性化学习模块

构建学生学习个性化推荐模块的核心任务之一是准确分析学生的兴趣、特长、潜能，用完备且准确的属性标签对学生学习情况进行全覆盖，从而极大促进精准学生个性化学习模块推荐。根据数据源层抽取的数据并且结合已构建的学习画像，利用KNN与朴素贝叶斯模型形成推荐列表。根据已确立的标签存入数据训练样本集，每条数据记录都有其对应的属性及标签。当输入新的学生记录时，此时该条数据不具备标签，将新数据中的样本与该条记录最相似的数据进行比对，从而提取标签集，故可根据新建后的标签进行聚类分析。提取学生学习时的特征即上述不同标签下的子属性计算学生学习偏好与学习数据库中的学习资源之间的相似度，再运用KNN分类器，按照远近距离分配学习资源给不同的用户群，形成学习资源的个性化推荐。针对学习资源推荐，分类的任务即为特定学生寻找合适的学习资源，用准确率(Precision)和召回率(Recall)衡量推荐成效，准确率表示学生对该项学习资源感兴趣的概率，召回率为学生感兴趣的资源被成功推荐的概率，准确率和召回率值越大表示推荐效果越好。用F表示准确率和召回率的调和平均值，其值越大表示推荐质量越高。

具体计算模型如下：

$$Precision = \frac{his_s}{recset_s} \tag{1}$$

$$Recall = \frac{his_s}{testset_s} \tag{2}$$

$$F = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

上式中， $his_s$ 表示成功推荐给学生的有效学习资源数量， $recset_s$ 表示推荐学习资源数量， $testset_s$ 表示符合学生需求的推荐学习资源数量，Precision代表准确率，Recall代表召回率。

召回步骤完成初筛，帮助分析学生学习兴趣偏好，为进入下一流程进行粗排和精排做准备。对学生学习、消费、网络使用及生活等行为数据进行分析，完成打分，从而最终推断出学生大致的学习风格，达到为学生推荐个性化学习资源的目的。

### 3.3 问题预警模块

根据已构建的学习画像，结合学生在校线上及线下统计数据建模，对学生课堂学习、上网信息、门禁记录等结果进行量化分析。运用BP神经网络、RBF径向基模型，输入相应向量训练网络以达到局部逼近任意连续函数<sup>[5]</sup>。考虑到在训练过程中分布逐渐偏移变动降低收敛速度，为防止模型过分拟合，故添加Batch Normalization层，为的是将输入的学生成绩数据数值进行标准化，缓解后期DNN训练中的梯度消失问题，加快模型的训练速度，使输出的特征图均匀度提升，增大梯度，提升收敛度，让模型趋于稳定，从而根据学生个人属性综合趋势对成绩稳定性和挂科率进行预测。分析学生学习效率与挂科率、网络使用、消费情况及失联记录等之间的关系，进而设立预警条件，达到预警目的。

### 3.4 辅助学校决策模块

学生画像的构建，可重点结合学校管理实际需求，分析所关联的学生数据。可以进行问卷调查，从而完成描述性统计，并结合上文所构建的学生学习画像模型<sup>[6]</sup>进行比对，直至提出最有利于学生的有关决策方案，为学校实现浅层干预与深层干预相结合的目标提供支撑，使制度政策能更好地服务于学生。

## 4 结论(Conclusion)

构建大学生学习画像，建立合理有效的数据挖掘模型，根据模型输出结果对学生进行个性化指导，具有一定的针对性和可操作性，并对改善学生学习效果、提高学生培养质量、发现有潜质的人才、提高学生综合素质具有重要的现实意义。大学生学习画像为高校管理者、教师提供了参考，有助于引导学生全面发展并发挥特长，为经济社会输送高质量、专业能力强的<sup>[7]</sup>人才。

## 参考文献(References)

- [1] 钱爱娟,董笑菊,沈绮文,等.高校图书馆用户画像与行为可视化分析[J].图书馆杂志,2020,39(10):82-88.
- [2] 魏孔鹏,谷洪彬,李啸龙,等.学生综合素质评价的用户画像构建研究[J].计算机时代,2020(03):96-98.
- [3] 吕拙拙.智慧校园视域下高校用户画像探究[J].大众标准化,2020(19):45-48.
- [4] 杨光莹,杜敏,杨东梅,等.基于校园行为数据分析的学生画像系统初步构建研究[J].教育教学论坛,2020(41):44-45.
- [5] 张丽,吕康银.智慧城市公共服务数据画像及应用模式研究[J].情报科学,2020,38(10):61-67;89.
- [6] 金冈增,李娜,郑建兵,等.科研人员画像系统设计与实现[J].软件工程,2018,21(09):28;41-43.
- [7] Ye Sun, Rongqian Chai. An Early-Warning Model for Online Learners Based on User Portrait[J]. Ingénierie des Systèmes' Information, 2020, 25(4):26-43.

## 作者简介:

陈会(1999-),男,本科生.研究领域:智慧校园.  
 余馨(2000-),女,本科生.研究领域:数据分析.  
 李琳琳(1999-),女,本科生.研究领域:信息经济.  
 吴苏徽(1998-),女,本科生.研究领域:数据分析.  
 蒋秀莲(1968-),女,本科,高级工程师.研究领域:信息管理,信息经济.