

在线教育平台中个性化学习资源推荐系统设计

李乡儒¹, 梁惠雯², 冯隽怡², 肖江平², 彭婉芬³

(1. 华南师范大学 计算机科学学院, 广东 广州 510631;

2. 华南师范大学 数学科学学院, 广东 广州 510631;

3. 广州犀灵信息科技有限责任公司, 广东 广州 511458)

摘要:为了通过充分挖掘和分析用户的学习行为规律及认知特点,借助互联网和人工智能技术提升个性化教育的深度和广度,设计了一个包含用户画像的个性化学习资源推荐系统。该系统由数据层、数据分析层和推荐计算层构成。数据层由用户数据以及包含知识资料、学习资料和标签集的资源库组成;数据分析层融合了以基础信息、学习行为等为代表的静态数据和动态数据,据此为用户生成个性化画像、提供直观形象的学习反馈;推荐计算层则通过相似性分析和聚类算法发现用户的学习行为规律,使用 TF-IDF 方法挖掘用户的资源偏好,并据此给出个性化的学习建议。该系统已应用于一个以人工智能类课程为主的在线教育平台,为师生提供个性化画像、学习反馈与资料推荐的服务,当前处于第二个学期的试用阶段。

关键词:在线教育;个性化推荐系统;用户画像与反馈;学习风格;行为序列分析;资源偏好

中图分类号:TP311

文献标识码:A

文章编号:1673-629X(2021)02-0143-07

doi:10.3969/j.issn.1673-629X.2021.02.027

Design of Personalized Learning Resource Recommendation System for Online Education Platform

LI Xiang-ru¹, LIANG Hui-wen², FENG Jun-yi², XIAO Jiang-ping², PENG Wan-fen³

(1. School of Computer Science, South China Normal University, Guangzhou 510631, China;

2. School of Mathematical Sciences, South China Normal University, Guangzhou 510631, China;

3. Guangzhou Xiling Information Technology Co., Ltd., Guangzhou 511458, China)

Abstract:To improve the depth and breadth of personalized education by fully exploring and analyzing the learning behaviors and cognitive characteristics of users with the help of the Internet and artificial intelligence technology, a personalized learning resource recommendation system with user portrait is designed, which consists of data layer, data analysis layer and recommendation calculating layer. The data layer includes user information and a repository containing knowledge materials, learning materials and tag sets. The data analysis layer integrates static data and dynamic data represented by basic information and learning behavior, so as to generate personalized portrait and provide intuitive learning feedback for users. The recommendation calculating layer finds out the patterns of learning behavior for users by similarity analysis and cluster analysis, mines the resource preference of users by using the TF-IDF method, and gives personalized learning suggestions accordingly. This system has been used in an online platform for artificial intelligence education tested its capabilities of providing teachers and students with personalized portrait, learning feedback and resource recommendation for the second semester.

Key words:online education; personalized recommendation system; user portrait and feedback; learning style; behavior sequence analysis; resource preference

0 引言

随着大数据、云计算、移动互联等科学技术的不断发展和人们对教育需求的提升,以在线教育平台为主

的“互联网+教育”模式得到了广泛的推广和普及。截止2019年6月,国内在线教育用户规模达2.32亿并呈快速持续扩张的趋势。目前,大多数的在线教育平

收稿日期:2020-03-29

修回日期:2020-07-30

基金项目:2019年度国家级大学生创新创业训练计划基金资助项目(201910574044);国家自然科学基金项目(11973022);广东省自然科学基金(2020A1515010710)

作者简介:李乡儒(1972-),男,博士,教授,通讯作者,CCF会员(42118M),研究方向为机器学习、计算机视觉及其应用、互联网+教育。

台以整合优质网络课程和其他学习资源为用户提供服务。然而,面对海量的学习资源,用户很难快速地找到适合自己的学习资料,甚至可能出现“认知负荷”、“信息迷航”等问题。因此,能够根据用户的学习过程智能化分析其学习行为规律及认知特点,并提供个性化的学习资源推荐是当前在线教育平台的一个迫切需求。

大多数在线教育平台的个性化推荐都是围绕着用户的特征信息进行分析和建模。例如,通过用户的学习风格、学习兴趣等个性化特征信息构建推荐模型^[1-2]。然而,这种推荐的结果往往比较粗糙,很难令人满意,一般适合于一些简单推荐。近些年,有研究开始尝试利用用户的行为特征信息去分析用户之间的相似度。例如,桂忠艳等通过计算用户的学习行为序列

相似度,采用基于用户的协同过滤推荐建模^[3-4]。然而,这些研究更多的是聚焦在用户的特征信息,忽视了对学习资源的挖掘分析。

该文设计了一个基于学习风格、资源偏好和行为序列的个性化学习资源混合推荐系统,在充分利用用户行为特征信息的前提下,结合学习风格和资源偏好,为用户生成个性化画像,进一步为用户提供学习反馈和个性化的学习资料推荐服务。

1 系统体系架构

设计的个性化学习资源推荐系统自下而上分为数据层、数据分析层和推荐计算层,应用于一个以人工智能类课程为主的在线教育平台的体系架构,如图1所示。

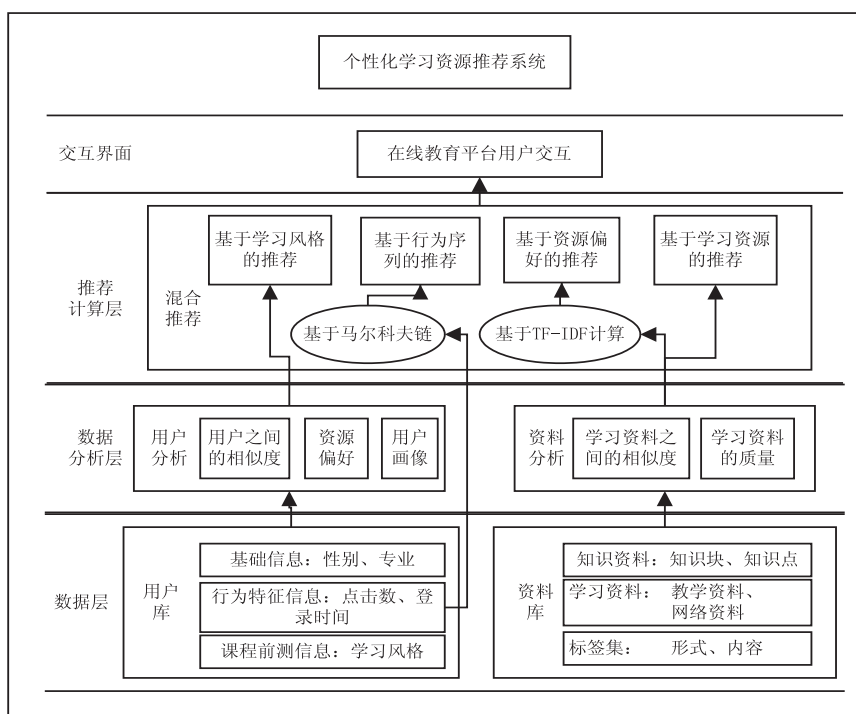


图1 个性化学习资源推荐系统的体系架构

1.1 数据层

1.1.1 用户库

用户库存储用户的特征信息,包括个性化特征信息和行为特征信息^[5]。个性化特征信息是指用户自身所固有的,且不随时间发生变化或变化缓慢的信息。典型的个性化特征信息包括用户的年龄、性别、专业等基础信息,如学习风格等课程前测信息。个性化特征信息在数据形式的表现上为静态数据。行为特征信息是指随时间推移有显著变化的信息。例如,登录时间、点击次数、发帖等。这类信息在数据形式的表现上为动态数据。在该研究中,根据用户以下三种学习方式,对行为特征信息进行分类:

(1) 自主学习,如观看视频、浏览学习资料等;

(2) 反思学习,如提交作业、查看作业批改结果等;

(3) 交流反馈,如讨论区留言等。

1.1.2 资源库

资源库由知识资料、学习资料和标签资料组成。知识资料包括知识块和知识点。在所实施的《模式识别》案例中,以章节的划分及中小标题作为知识块,以各章节内容所出现的定理、算法、数理名词等作为知识点。学习资料是指以知识资料中的知识块、知识点作为关键词进行网络爬取,并做一定的人工筛选后的内容资料。

标签集是指系统中用于概述学习资料的内容与形式的概括性描述。标签不仅简洁、直观地概述了学习

资料,便于用户快速浏览和选择,而且能够将其转化为相对应的文本字段^[6],便于数据的深度挖掘和分析。在所实施的《模式识别》案例中,从内容和形式两个方面设计了 7 类标签,具体如下:

(1)内容:知识块(标签值:中标题、小标题);知识点(标签值:算法、定理、梳理名词)。

(2)形式:语言(标签值:中文、英文);类别(标签值:推导、实现、总结、推广);载体(标签值:文字、图片、视频);数据集(标签值:MNIST、sklearn data、其他);编程语言(标签值:java、c++、python、其他)。

1.2 数据分析层

1.2.1 用户分析

系统通过对用户的个性化特征信息和行为特征信息的量化、统计和建模,对其进行挖掘分析,包括用户之间的相似度分析、资源偏好分析和用户画像。

用户之间的相似度分析是进行用户推荐建模的基础。它是通过用户的特征信息来计算用户之间的相关程度,从而确定用户之间的相似度,并把相似用户称为“邻居用户”,然后将“邻居用户”所选择的学习资源推荐给当前用户。

资源偏好是指用户对学习资源的内容和形式的偏爱、喜好等倾向性要求。例如,有的用户喜欢以文本为载体的学习资源,而有的用户则喜欢视频类型的学习资源。在该系统中,使用 TF-IDF 算法^[7]计算用户下每一类标签的标签值所占的权重,从而可得出用户的资源偏好。

用户画像是建立在一系列真实数据之上的用户模型。它可以从多角度描述用户的学习特点。与大多数仅用个性化特征信息构建用户画像的在线教育平台不

同,该系统既结合了用户的个性化特征信息,也考虑了行为特征信息,定量和定性地构建了用户的个性化画像^[8-9]。例如,通过用户的课程前测信息得出学习风格;通过对用户的浏览时长、浏览学习资料的频率等行为特征信息统计得出其学习的常用模块序列、知识兴趣点、知识难点、作业详情等。用户可以通过个性化画像了解和掌握自己的学习情况,便于调整学习策略。

1.2.2 资料分析

将标签、点击数(阅读量)、评论数等作为学习资料的属性特征。系统通过对学习资料属性特征的量化、统计和建模,进行学习资料的相似度分析和质量分析等。

学习资料之间的相似度分析是基于学习资源推荐的建模基础。它是将学习资料的标签作为特征来计算学习资料之间的相关程度,从而确定学习资料之间的相似度,并把相似资料称为“邻居资料”,然后将“邻居资料”推荐给当前用户。

学习资料的质量分析主要是通过对点击数(阅读量)、评论数等属性特征的统计分析,可在一定程度上过滤掉劣质内容。

1.3 推荐计算层

1.3.1 基于学习风格的推荐

在线学习平台通过引导用户完成具有较好可信度的课前学习测评任务,分析得出用户的学习风格。进而,系统根据用户的学习风格进行相应的学习资源推荐。该文采用 Felder-Silverman 模型^[10]的 4 个维度对用户的学习风格进行划分,分别是信息加工、信息感知、信息输入和信息理解,具体见表 1。

表 1 学习风格显式分类

风格维度	维度分类	特征	相关推荐
信息加工	活跃型	更倾向通过实验验证想法,先做后想,更喜欢讨论、团队合作	例如:讨论区,论坛
	沉思型	更倾向通过思考学习新事物,先想后做,更喜欢独立思考、工作	例如:博客、日志
信息感知	感悟型	更倾向实例、记忆	更多具体的实例,例如算法应用
	直觉型	更倾向抽象、推导	更多抽象的推导,例如算法延伸
信息输入	言语型	更倾向解释性文字	PPT,文献
	视觉型	更倾向图片视频等	视频、思维导图
信息理解	综合型	先大体理解再深入	更多“上下节”标注的课程或知识导航
	序列型	一步步按先后学习	更多大体提纲概要

1.3.2 基于行为序列的推荐

基于用户的个性化特征信息的推荐虽然被广泛应用,但是这种推荐往往比较粗糙。因此,这类方法一般适合于简单推荐,而对于复杂一些的推荐任务则很难

得到令人满意的结果。近些年,有研究开始尝试利用用户的行为特征信息去分析用户之间的相似度。该系统以桂忠艳等^[4,11-12]对网络学习行为数据的研究为基础,对用户相似度计算的稳定性进行改进。

1.3.2.1 相似度计算

(1) 行为序列相似度的计算。

用户的行为序列用有限集 S 表示:

$$\{(z_1, y_1), (z_2, y_2), \dots, (z_n, y_n)\}, n \geq 2$$

其中, (z_i, y_i) 表示第 i 个元素对, z_i 表示访问模块, y_i 表示相应操作, 按照行为发生的先后顺序依次记录在集合中。为了描述的简洁, 将元素对 (z_i, y_i) 中的元素连接组成的字符串 $z_i y_i$ 用 s_i 表示, 称之为用户的第 i 个状态串。

状态序列是行为序列中每个元素对中的元素依次链接形成的字符串。例如, 用户 S 的状态序列用状态串可表示为“ $s_1 s_2 \dots s_n$ ”, 记为 $S = s_1 s_2 \dots s_n$ 。状态序列 S 的状态子序列定义为 $S^{(i)} = s_{n_1} s_{n_2} \dots s_{n_i}$, 其中, $1 < n_1 < n_2 < \dots < n_i < s$ 。

令用户 A 和 B 的状态序列分别为 A 和 B , 则行为序列的相似度^[4]由以下计算公式给出。

$$\text{sim} = \alpha \times \text{sim}_{\text{seq}}(A, B) + \beta \times \text{sim}_{\text{trans}}(A, B) + \gamma \times \text{sim}_{\text{value}}(A, B)$$

其中, $\alpha + \beta + \gamma = 1, \alpha \geq 0, \beta \geq 0, \gamma \geq 0$ 。 $\text{sim}_{\text{seq}}(A, B) = \frac{i_{\max}}{s + l}$, $\text{sim}_{\text{trans}}(A, B) = \frac{\varphi}{(s - 1) + (l - 1)} \text{sim}_{\text{value}}(A, B)$ $= \frac{|A \cap B|}{|A \cup B|}$ 分别表示状态值相似度、状态转移相似度和状态次序相似度。

(2) 基于时间衰减效应的用户相似度计算。

用户在不同时间段的学习行为对预测其学习行为具有不同的贡献。一般来说, 发生时间越靠近的行为更能体现用户的学习兴趣, 对用户之间相似度的贡献也越大。为提高近期的行为序列对相似度计算的重要性, 引入时间权重函数 WT 。

$$WT(A, S_i) = (1 - a) + a \frac{D_{A, S_i}}{L_A}$$

其中, S_A 为用户 A 所有的行为序列集, D_{A, S_i} 表示用户 A 生成的行为序列与其最早生成的某行为序列的时间间隔; L_A 表示用户 A 行为序列的时间跨度; $\alpha \in (0, 1)$ 为权重增长指数。故用户 A 和 B 之间的基于时间衰减效应的用户相似度计算如下:

$$\text{sim}(A, B) = \frac{\sum_{S_i \in S_A} [(W(A, S_i) + W(B, S_j))/2] \cdot \text{sim}(S_i, S_j)}{|S_A| \cdot |S_B|}$$

在分析用户之间的关系时, 仅考虑行为相似度是不够的。造成相似度极高的原因有很多, 例如, 在较短时间内不能观察到用户长期行为的差异。然而, 在实际应用中需要对用户之间的关系进行更精准、更稳定的描述^[13]。为此, 该文提出相关系数的概念, 即通过分析一段时间内相似度的变化, 得到该时间段内用户

之间的相似度。假设平均相似度为 sim_{avg} , 方差为 sim_{dx} , 则相关系数 (RC) 可通过下式计算:

$$RC = \frac{\text{sim}_{\text{avg}}}{\text{sim}_{\text{dx}}}$$

因此, 两个用户的关系越相近, 平均相似度变化越大; 反之, 平均相似度变化越小。

1.3.2.2 基于行为序列的推荐原理

为了解决传统的最近邻协同过滤推荐由于用户规模扩大导致搜索的时效降低难以满足系统的实时性需求问题^[14], 该系统首先采用 K -Means 算法对用户进行聚类, 然后在用户的聚类空间内计算行为序列相似度, 最后根据确定的“邻居用户”将其所选择的学习资源推荐给当前用户。

1.3.3 基于资源偏好的推荐

系统通过 TF-IDF 算法计算出用户下每一类标签的标签值的权重大小, 获得用户的资源偏好, 进而得到学习资源推荐的依据。例如, 对于喜欢文本类的用户, 系统更多地向其推荐文本载体的学习资料。

1.3.4 基于学习资源的推荐

系统通过标签计算学习资料之间的相似度^[15], 并把相似资料称为“邻居资料”, 然后将“邻居资料”推荐给当前用户。

1.3.5 混合推荐机制

由于每一种推荐方法都有其优缺点和适用特定的场景, 而且, 用户在进行线上学习的过程中, 也是一个不断变化的动态过程, 故仅仅考虑一种推荐方法进行学习资源推荐不一定符合实际的应用情形。由此, 提出了一种基于上述 4 种推荐方法的混合推荐机制。

在课程学习初期, 根据用户的课前学习测评获得用户的学习风格, 采用基于学习风格的方法推荐学习资料; 随着用户学习时间的增加和交互的增多, 对用户的学习行为进行挖掘分析, 采用基于学习行为的方法推荐学习资料; 同时, 可以通过用户的行为数据获得用户的资源偏好。采用了以基于资源偏好的推荐为主、学习资源推荐为辅的协同过滤推荐。该混合推荐机制如图 2 所示。

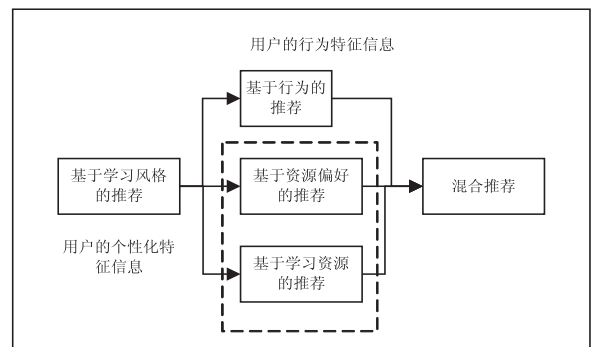


图 2 混合推荐机制

2 推荐系统业务流程

前述 1.3 节介绍了文中推荐算法的基本原理,这些算法基于对学习风格、资源偏好和行为序列的分析

可给出个性化学习资源混合推荐系统。该混合推荐系统的流程如图 3 所示。

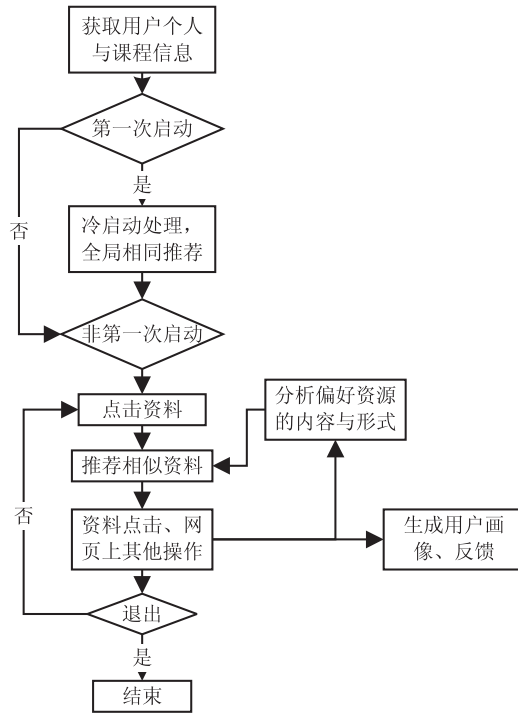


图 3 推荐系统业务流程

进一步描述如下:

第一步:用户登录。获取用户信息和课程信息。

第二步:判断用户是否初次学习该课程。若是,引导用户完成课前学习测评任务,分析得出用户的学习风格,根据用户的学习风格进行相应的学习资料推荐。

第三步:行为特征信息收集。系统采集行为特征信息的相关数据,如:登录时间、点击次数、发帖等。

第四步:分析用户的资源偏好、计算用户之间的相似度和学习资源之间的相似度,根据混合推荐机制向用户推荐学习资料。

第五步:判断用户是否退出系统。若否,则跳转至第三步。

范大学数科院所开设的《模式识别》和《数据挖掘导论》课程教学中。系统通过充分利用用户的学习行为数据,结合学习风格和资源偏好,为同学们推荐优质的学习资料。与此同时,向授课教师和学生反馈教学和学习情况。其中,图 4 展示了用户的课程信息,图 5 为系统作业详情及学习资源推荐内容。表 2 是对资源模块的说明。

图 6 为某一学生的用户画像与反馈,其中的三幅统计图表展示了作业分析情况。通过对该生的学习行为进行分析,发现其常用学习序列如下:

(1) 课件浏览-资料 A 点击-资料 B 点击-课程讨论区-作业查阅。

(2) 作业查阅-课件浏览-资料点击-课件浏览-资料提交-作业提交。

3 系统应用

文中的个性化学习资源推荐系统已应用于华南师

课程管理		模式识别2019	
我的课程 加入新课程		课程名	模式识别2019
课程名	课程描述	班级名	信息与计算2017
贝叶斯学习	数据挖掘研讨班课程	授课老师	李乡儒
数据挖掘导论	2018-2019-2	课程描述	2019-2020-1
模式识别2019	2019-2020-1	课程作业	点击此处查看所有作业
		课件	点击此处查看所有课件
		教学视频	点击此处查看所有教学视频
		课程讨论区*	点击此处进入课程讨论区

图 4 课程信息

模式识别2019		推荐内容
Softmax_Regression作业		特征选择理解及代表性算法解析【中文】
作业标题	Softmax_Regression作业	LASSO与岭回归[文段][中文][实现][Python]
作业详情	Softmax_Regression作业, PDF文件格式	相关系数 正则化[文段][中文][实现][Python]
开始日期	2019年9月29日 00:00	特征子集搜索 特征子集评价函数(熵)[文段][中文][推导]
截止日期	2019年10月9日 23:59	特征子集搜索 特征子集评价函数(信息增益)[文段][中文][推导]
本次作业包含的代码文件:		贪心策略[文段][中文][实现]
本次作业代码使用的数据集		ReliefF算法[文段][中文][实现][Matlab][其他数据集(UCI)]
本次作业代码运行结果	尚未开始批改作业	熵 信息增益[文段][中文][实现][其他数据集(stand ford-nlp)]
本次代码作业的成绩	暂无成绩	相关系数 特征排序[文段][中文][实现][Python]
本次PDF作业的成绩		
提交pdf	命名格式为: 自己学号_队友学号_自己名字_队友名字 或 自己学号_自己名字 选择文件 未选择任何文件 提交 未提交pdf文件	
已提交的pdf作业		

图 5 作业详情及学习资源推荐内容

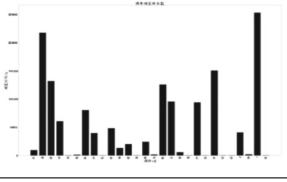
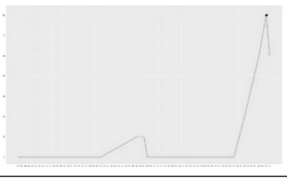
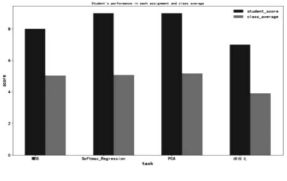
姓名	李明		年龄	19	班级	17 信计			
基本信息	在学课程	模式识别	作业分析	各次课件总浏览时长 其中横轴表示课件名称, 纵轴表示时长(单位: 秒)。					
	已学相关课程	数据挖掘 数学分析 高等代数							
学习风格	信息加工	沉思型					学习者作业提交情况 横轴表示时间(单位: 天), 纵轴表示作业提交的份数。		
	信息感知	感悟型							
	信息输入	视觉型							
	信息理解	综合型							
常用学习模块序列	作业查阅-课件浏览-资料点击-课件浏览-资料提交-作业提交						学习者成绩情况反馈图 横轴表示科目, 纵轴表示时间, 蓝色表示学生个人的成绩, 橙色表示班级的平均成绩。		
	课件浏览-资料 A 点击-资料 B 点击-课程讨论区-作业查阅								
兴趣知识块/点	短期	Relief-F 算法 贪心策略							
	长期	特征选择模型							
难点知识块/点	短期	LASSO 与岭回归							
	长期	正则化							

图 6 用户画像

表 2 资源模块说明

资源/模块	编码	动作	编码	说明
登录页面	S ₀	登录	a ₁	输入个人信息登录
我的课程	S ₁	查看	a ₂	个人课程中心, 包含所选课程
课程页面	S ₂	查看	a ₂	包含课程名、班级名、授课老师、课程描述, 以及课程作业、课件、教学视频和课程讨论区的链接
视频页面	S ₃	观看	a ₃	某课程下的教学视频界面
课程讨论区	S ₄	讨论	a ₄	问题讨论区, 可查阅问题和提问
助课页面	S ₅	浏览	a ₅	学生上传的代码
课件目录页面	S ₆	浏览	a ₅	包含教师课件, 助课课件的目录
课件页面	S ₇	浏览	a ₅	老师上传的课件
课件难度按钮	S ₈	点击	a ₆	笔记分类难点, 对应知识块知识点
推荐资料	S ₉	点击	a ₆	包含点击链接的标签(形式)和知识块知识点
作业页面	S ₁₀	查看	a ₂	包含课程布置的所有作业
某作业页面	S ₁₁	查看	a ₂	某一次作业
作业	S ₁₂	提交	a ₇	某一次作业提交
网站说明	S ₁₃	查看	a ₂	网站说明

学生从常用学习模块序列的反馈中可以知晓自己的学习路径。这有助于学生进一步了解自己的学习过程和学习行为规律,也方便授课教师了解学生的学习特点。

4 结束语

设计了一个应用于人工智能类课程的个性化学习资源推荐系统,该系统包括数据层、数据分析层和推荐算法层。系统在充分利用用户学习行为数据的前提下,结合学习风格和资源偏好,综合4种推荐方法制定了混合推荐机制为用户提供个性化画像、学习反馈与学习资料推荐的服务,当前处于第二个学期的试用阶段。

参考文献:

- [1] OCEPEK U, BOSNIĆ Z, ŠERBEC I N, et al. Exploring the relation between learning style models and preferred multimedia types [J]. *Computers & Education*, 2013, 69: 343–355.
- [2] MAMPADI F, CHEN S Y, GHINEA G, et al. Design of adaptive hypermedia learning systems: A cognitive style approach [J]. *Computers & Education*, 2011, 56(4): 1003–1011.
- [3] 吴淑苹. 基于数据挖掘的教师网络学习行为分析与研究 [J]. *教师教育研究*, 2013, 25(3): 47–55.
- [4] 桂忠艳, 张艳明, 李巍巍. 基于行为序列分析的学习资源推荐算法研究 [J]. *计算机应用研究*, 2020, 37(7): 1979–1982.
- [5] 陈佳艳. 基于学习行为特征的学习资源个性化推荐研究 [D]. 南京: 南京邮电大学, 2018.
- [6] 谢意, 陈德人, 干红华. 基于浏览偏好挖掘的实时商品推荐方法 [J]. *计算机应用*, 2011, 31(1): 89–92.
- [7] SALTON G, BUCKLEY C. Term-weighting approaches in automatic text retrieval [J]. *Information Processing & Management*, 1988, 24(5): 513–523.
- [8] 亓丛, 吴俊. 用户画像概念溯源与应用场景研究 [J]. *重庆交通大学学报: 社会科学版*, 2017, 17(5): 82–87.
- [9] 王庆, 赵发珍. 基于“用户画像”的图书馆资源推荐模式设计与分析 [J]. *现代情报*, 2018, 38(3): 105–109.
- [10] 裴艳. 基于学习分析的学习资源个性化推荐研究 [D]. 西安: 陕西师范大学, 2015.
- [11] 谭明新, 鲍晓琴. 基于用户模型的生成性学习资源个性化推荐研究 [J]. *软件导刊*, 2017, 16(8): 123–127.
- [12] MANAVOGLU E, PAVLOV D, GILES C L. Probabilistic user behavior models [C] // Third IEEE international conference on data mining. Piscatawy, NJ: IEEE, 2003: 203–210.
- [13] 关海潮. 一种基于行为序列的网络用户异常行为分析方法 [D]. 南京: 南京邮电大学, 2018.
- [14] 王辉, 高利军, 王听忠. 个性化服务中基于用户聚类的协同过滤推荐 [J]. *计算机应用*, 2007, 27(5): 1225–1227.
- [15] 熊楚平. 基于标签权重的个性化协同过滤推荐算法 [D]. 乌鲁木齐: 新疆大学, 2019.