

Development and Design of Mobile Intelligent Learning Platform on Collaborative Filtering

Kaixin Yang, Yawei Li

College of Software, Tianjin Normal University, Tianjin
Email: 13110012630@163.com, llyywwaoao@163.com

Received: May 28th, 2019; accepted: June 12th, 2019; published: June 19th, 2019

Abstract

In the education app market, there are fewer comprehensive education apps only providing online course platform learning services. Some education classes only provide the status of teaching assistant functions. We have proposed a teaching aid for colleges and universities. This paper put forward a mobile intelligent learning platform for students to learn online courses. The platform introduces the KNNBaseline collaborative filtering algorithm, proposes to first calculate the similarity between the target user and the user preset by the system, and then perform similarity calculation with the remaining users, effectively solving the problem. The information is too sparse to cause problems, and introduces a baseline model in which the user's score changes over time. Experiments were carried out using the MovieLens ml-100k dataset. The experimental results show that the algorithm can effectively provide personalized recommendation services for students.

Keywords

Collaborative Filtering, KNNBaseline, Scikit-Surprise, Online Class Learning, Educational Administration, WeChat Applet

基于协同过滤算法的移动智能学习平台的开发与设计

杨凯欣, 李雅玮

天津师范大学软件学院, 天津

Email: 13110012630@163.com, llywwaoao@163.com

收稿日期: 2019年5月28日; 录用日期: 2019年6月12日; 发布日期: 2019年6月19日

摘要

针对教育类app市场中, 综合类教育类app较少并仅提供网课平台学习服务, 部分教育类仅提供教学辅助功能的现状, 本文提出了一款既能为高校提供教学辅助, 又能服务学生进行网课学习的移动智能学习平台, 该平台引入了KNNBaseline协同过滤算法, 先将目标用户与系统预设的用户进行相似度计算, 然后再与剩余用户进行相似度计算, 有效解决因为用户信息过于稀疏可能造成的问题, 并且引入了用户评分随着时间改变而改变的基线模型。使用MovieLens ml-100k数据集进行实验, 实验结果表明, 该算法可以有效地为学生提供个性化推荐服务。

关键词

协同过滤, KNNBaseline, Scikit-Surprise, 网课学习, 教务管理, 微信小程序

Copyright © 2019 by authors and Hans Publishers Inc.

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

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



Open Access

1. 引言

1.1. 项目开发背景

通过极光大数据可得, 截止 2017 年 8 月, 在线教育 app 市场规模达到 2.76 亿, 行业整体渗透率为 29.3%, 而随着国人自我增值需求的不断加强, 这个市场仍然有巨大的发展潜力。根据极光大数据统计, 截至 2017 年 8 月最后一周, 渗透率 top3 的综合教育类 app 分别是网易公开课、腾讯课堂和中国大学 MOOC, 其市场渗透率分别为 0.46%、0.30% 和 0.20% [1]。这些综合类教育 app 仅提供网课平台学习服务, 大多与高校教学脱节。

与此同时, 根据腾讯 2018 年一季报数据, 微信及 WeChat 合并 MAU 达到 10.4 亿, 超过 2017 年底我国 7.53 亿的手机网民规模, 微信已实现对国内移动互联网用户的大面积覆盖。2017 年微信登录人数已达 9.02 亿, 较 2016 年增长 17%, 日均发送微信次数为 380 亿, 微信已成为全民级移动通讯工具。微信已成为国内最大的移动流量平台之一。根据智研咨询集团提供的微信小程序 2017 年 12 月 Top200 的行业分布, 教育占比仅有 1%, 且多为工具性质[2]。

1.2. 项目开发内容及目的

根据综合类教育 app 的现状和微信便捷高效的特点, 本团队使用微信开发者平台, 设计并开发了一款以提供一体化教学服务为目标的教学辅助平台。该平台不仅可以为学生提供与高校紧密结合的教学服务, 高校教师还可以通过此平台上传相关教学课程, 学生还可以根据自身需求利用此平台进行网课学习。此外, 通过基于用户的协同过滤算法, 本项目还为平台设计并训练了一个推荐系统, 能够根据学生的行为信息为学生个性化推荐课程, 充分实现了学生的个性化教育。该系统还可以统计在校学生和教师的各

种数据, 为高校的教育提供准确、及时的学生信息, 同时也为高校对学生的教育提供引导信息。

2. 系统分析与设计

2.1. 系统功能设计

基于微信的智能网课学习平台为不同用户提供不同的入口。管理员通过后台管理端进行学习平台信息管理, 教师通过教师端进行教学管理, 学生通过学生端进行课程学习。具体功能模块如图 1 所示。

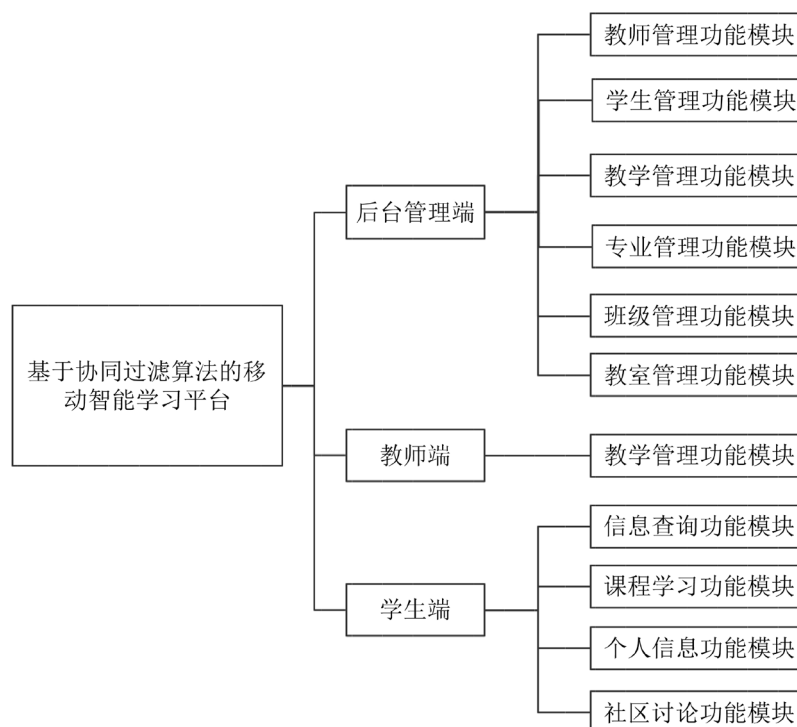


Figure 1. The functional module diagram of system [3]
图 1. 系统功能模块图[3]

2.2. 系统架构

系统有由三个子系统组成。分别为后台子系统, 课程推荐子系统, 用户客户端子系统。

用户通过用户客户端子系统发出请求数据, 后台子系统处理请求数据并返回用户想要的的数据, 并通过多样的形式将数据呈现在浏览器或微信小程序上。随着用户访问该平台次数的增加, 产生的用户日志也随之增加。课程推荐子系统会在线下根据用户日志分析用户行为信息, 计算用户的课程推荐列表, 并将用户的课程推荐列表数据传输到后台子系统, 后台子系统对推荐列表数据进一步处理后呈现到用户客户端。

1) 后台子系统

后台子系统从下往上依次由底层数据库, 数据模型层, 数据访问层, 业务逻辑层, 控制层组成。数据库负责存储数据; 数据模型层实现对象关系映射(ORM); 数据访问层对数据库进行增删改查; 业务逻辑层处理应用的业务逻辑和业务校验; 控制层处理用户请求数据, 实现业务功能。

后台子系统为用户和后台数据库的交互提供支持, 从而为不同用户提供不同的功能和服务。同时记录用户的行为信息并生成用户日志。

2) 课程推荐子系统

课程推荐子系统由应用层, 云平台计算层, 终端感知层组成。

根据后台子系统提供的用户日志, 课程推荐子系统对用户行为信息进行预处理, 在终端感知层量化用户行为信息; 在云平台计算层运用协同过滤算法得出推荐列表; 在应用层预测用户行为, 优化推荐列表。为用户推荐个性化课程。

3) 用户客户端子系统

分别采用 B/S 架构实现后台管理。管理员使用浏览器登陆后台管理网站完成各项后台管理操作。采用 C/S 架构为教师和学生提供界面交互和平台服务。教师和学生使用微信小程序, 通过授权访问学习平台。

系统架构图如图 2 所示:

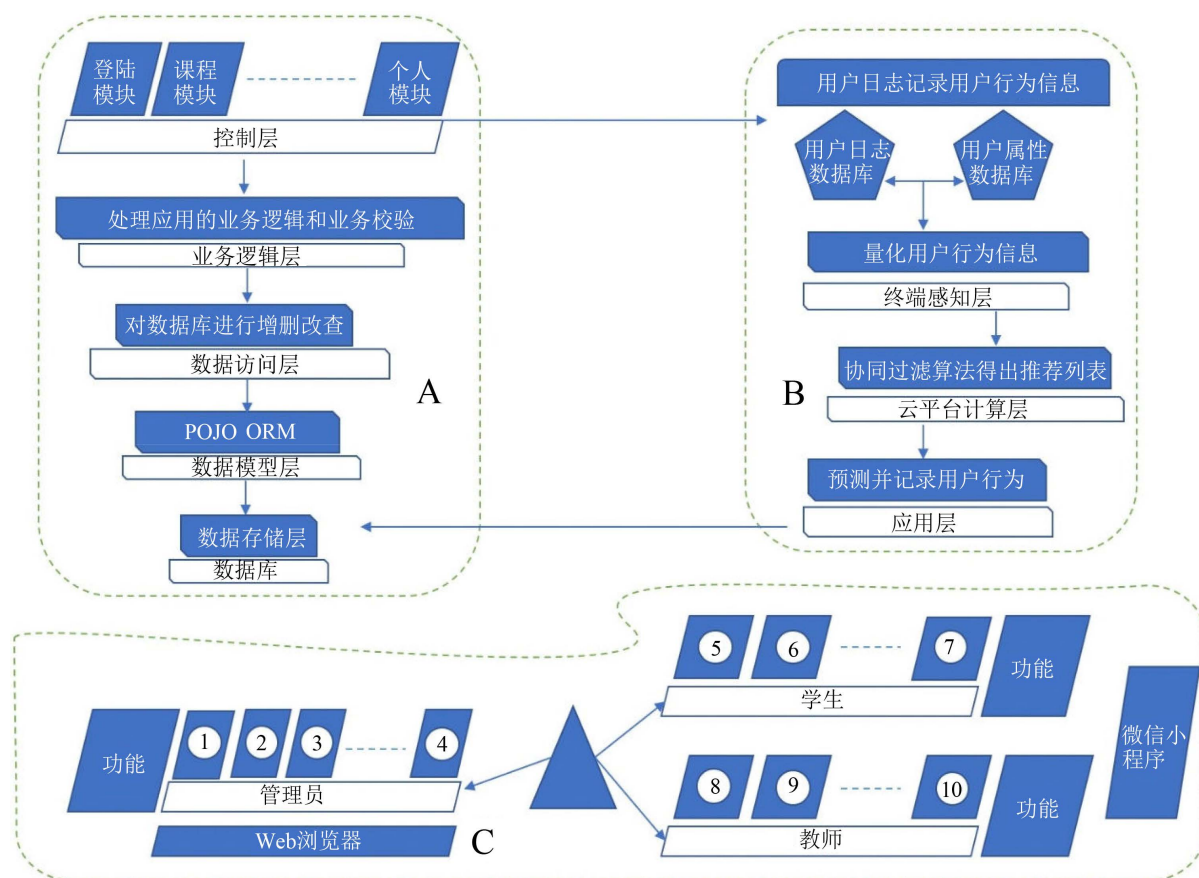


Figure 2. System architecture diagram [4]

图 2. 系统架构图[4]

图 2 中字母 A 代表后台子系统; 字母 B 代表课程推荐子系统; 字母 C 代表用户客户端子系统; 数字 1 代表学生管理; 数字 2 代表教师管理; 数字 3 代表教学管理; 数字 4 代表权限管理; 数字 5 代表个人信息; 数字 6 代表课程学习; 数字 7 代表信息查询; 数字 8 代表回答问题; 数字 9 代表发布通知; 数字 10 代表学生考勤。

2.3. 数据流图

数据流程是数据在系统中产生、传输、加工处理、使用、存储的过程[5]。

数据流程分析即是系统功能模块详细设计的基础,也是数据库设计的基础。开发者可以通过分析数据流明确数据传输和处理过程。该系统的数据流图分为顶层,一层和二层。顶层数据流图如图3所示[6]。

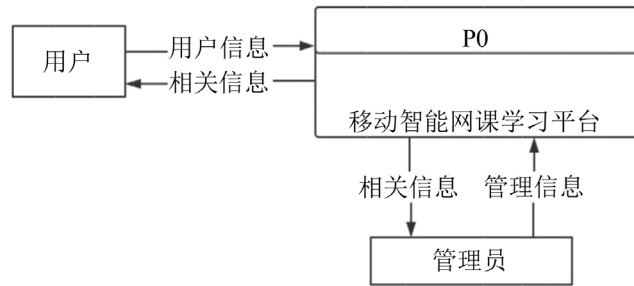


Figure 3. Top-level data flow diagram [7]
图3. 顶层数据流图[7]

将处理 P0 展开, 得到第一层数据流图如图4所示。

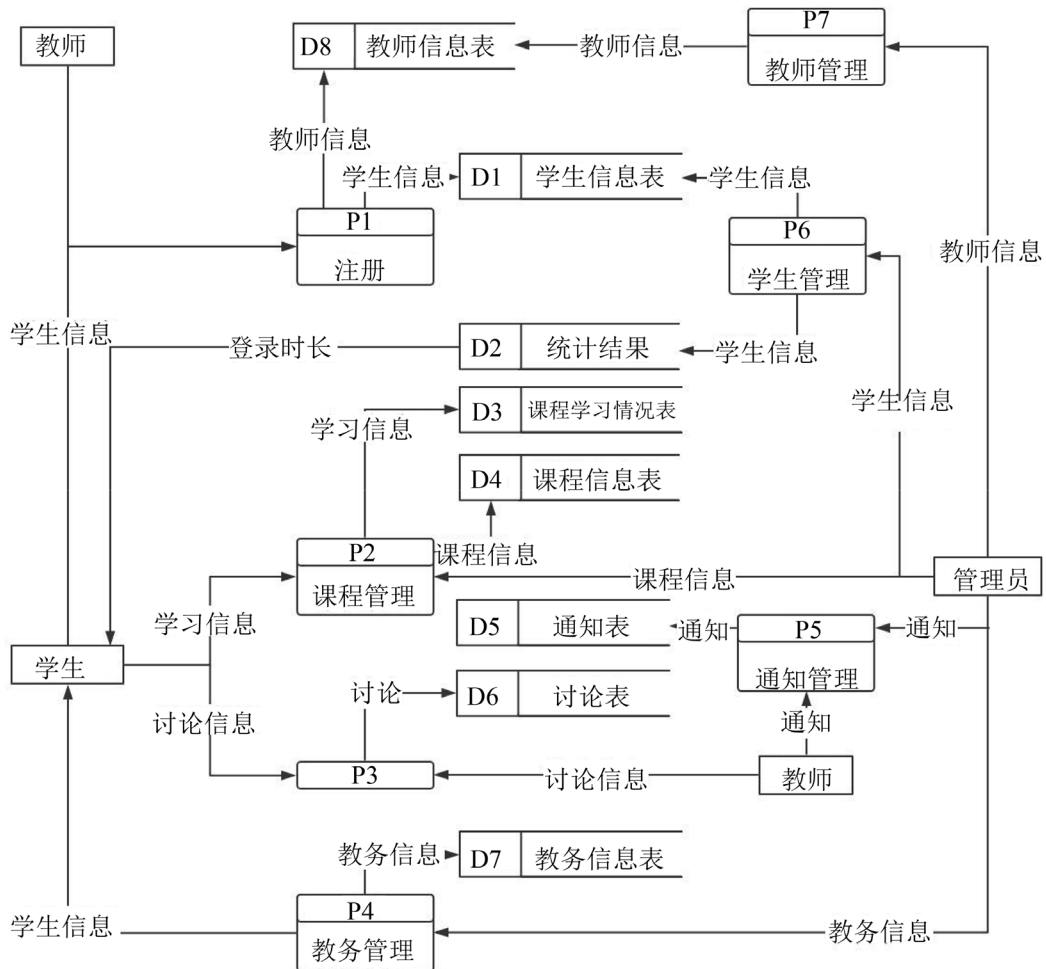


Figure 4. First-level data flow diagram [7]
图4. 第一层数据流图[7]

二层以 P5 为例, 将 P5 展开得到二层数据流图, 如图5所示。

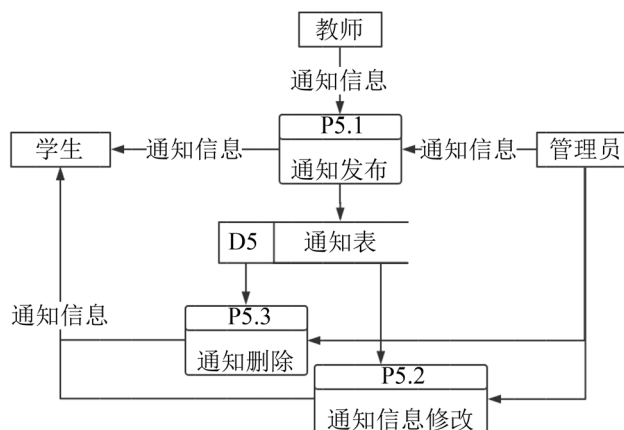


Figure 5. Second-level data flow diagram [7]
图 5. P5 展开数据流图[7]

3. 系统数据库设计

3.1. 数据库概念设计

数据库是一个系统的基础，是系统数据的源头，所以系统数据库的设计至关重要。通过需求分析得出，学生，教师，课程，专业，学院，通知等实体以及对应属性。根据实体之间的依赖关系，设计系统概念模型，遵循了完整性约束，三范式原则[6]。系统 E-R 图如图 6 所示。

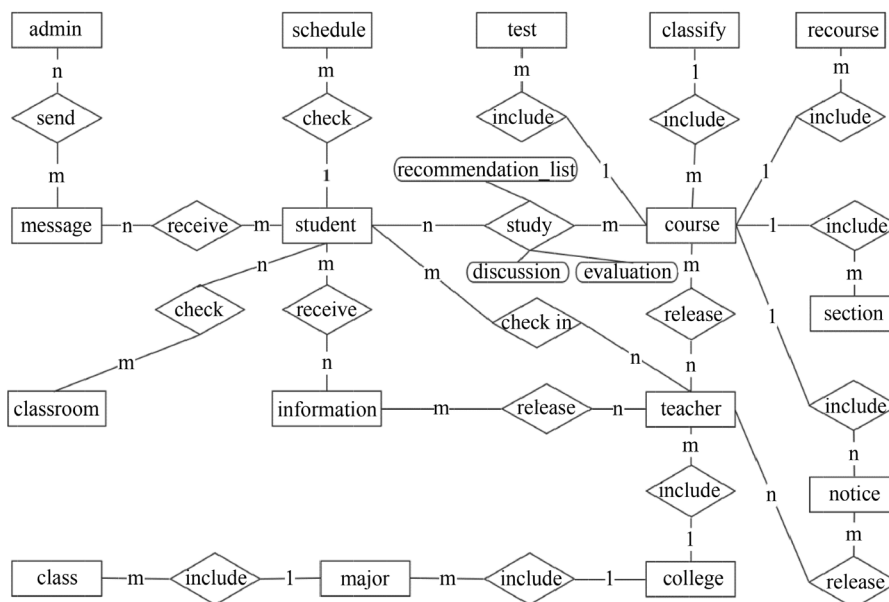


Figure 6. E-R diagram of system [7]
图 6. 系统 E-R 图[7]

3.2. 数据库逻辑结构设计

根据概念模型 E-R 图对数据库实现具体表的详细设计[8]。

以用户信息表为例。用户信息表保存了用户的基本信息，包括其所在学校信息和作为微信用户的信息，并标记用户身份类型(学生，教师，管理员)。

字段名, 数据类型, 数据长度等详细信息见表 1。

Table 1. The information of user table

表 1. 用户信息表

Field name	Data Type	Not Null	Remark
user_id	varchar32	NO	student id
user_name	varchar50	NO	student name
password	varchar32	NO	password
open_id	varchar128	NO	open id
nickname	varchar50	NO	nickname
gender	tinyint1	YES	gender, default man
header	varchar512	NO	header picture url
status	tinyint1	NO	0-unauthorized, 1-authorized
major_id	varchar32	YES	major id
class_id	varchar32	YES	class id
college_id	varchar32	YES	College id
identity	tinyint3	NO	0-student, 1-teacher
create_time	timestamp	NO	record creat time
update_time	timestamp	NO	record update time
del	tinyint1	NO	default not delete record

4. 系统开发

4.1. 系统开发环境

基于协同过滤算法的移动智能学习平台的开发环境为 Windows10, 64 位操作系统。软件使用 IDEA 集成环境和微信小程序开发平台。开发语言为 Java、JavaScript, CSS。MySQL5.7.19 作为后台数据库, 使用 InnoDB 存储引擎。

4.2. 开发技术

4.2.1. 客户端技术方案

微信小程序开发使用官方 MINA 框架编写, 前端程序自主开发, 贴近核心代码。开发过程中, 将对于大量相似却不重复的功能样式, 独立开发出区块接口, 划分出网络通信, 数据储存, 数据处理等模块 [9]。模块涵盖多种功能, 扩展性强, 二次开发性好。文件符合官网文件系统设定, 为不同功能区块, 相似 UI 设定开发出相应文件系统。

4.2.2. 服务器技术方案

服务器采用 J2EE, 该技术用于开发企业级应用。使用基于 SpringBoot 的 SSM 开发框架, 对后台进行高度封装。由数据存储层, 数据模型层, 数据访问层, 业务逻辑层和控制层组成。数据存储层存放系统的 MySQL 数据库; 数据访问层专注于数据库的增删改查操作; 业务逻辑层完成业务处理。控制层通过控制器实现前端和后台的数据序列化传输。同时使用 redis 缓存技术, 提高系统性能和效率。

5. 基于用户行为信息相似度的协同过滤算法

基于 scikit-surprise 的协同过滤算法

SciPy 是一个开源的基于 Python 的科学计算工具包。Scikit-surprise, 是基于 SciPy 针对机器学习推荐系统领域发展出的一个简便的推荐系统引擎。本系统使用了协同过滤算法, 在传统的协同过滤算法上, 引入了一个随时间变化的模型来刻画用户的评分之间的相似度[10]。传统的协同过滤算法, 简单来说是利用用户之间的兴趣相投, 共同喜好为用户推荐其感兴趣的信息[11], 其基本思想是利用用户对各类课程进行评分获得的矩阵, 在用户群中寻找高度相似的用户群, 然后综合这些用户对同一信息的评价, 预测目标用户对此信息的喜好程度[11]。协同过滤的实现一般分为两步: 首先, 获得用户对其所喜好课程的评价信息, 再构造评分矩阵, 当用户和课程种类的数量都很大时, 由于用户通常只对感兴趣的产品进行评分, 因此建立的评分矩阵是个巨大的稀疏矩阵 $Pm * n$, 见表 2。

Table 2. The information of user's course score table

表 2. 用户的课程评分表

Users	Higher Mathematics	Data Structures	Operating System	Computer Network
User 1	3	4	-	-
User 2	4	3	5	4
User 3	2	1	-	-
User 4	4	-	5	1

-代表用户未对此课程进行评分。

表 1 的一行的数据代表一个用户对不同电影的评分, 也标志了该用户的喜好程度。其次为了分析用户的相似性, 即通常情况下, 口味相近的人往往喜欢的课程也会比较相似, 因此要计算用户的相似性, 一般会补全评分矩阵来减少误差[11]。此处, 该课程推荐系统会优先考虑同一个学院的学生进行相似度推荐, 然后再筛选了同一门课的学生进行相似度推荐, 从而缩小评分矩阵, 降低其稀疏程度。并且发现因为用户行为信息在不断变化, 根据用户评级, 用户对对应课程的评分也会随之变化, 提出了 KNNBaseline Model。由 b_{ui} 表示

$$b_{ui} = \mu + b_u + b_i \quad (1)$$

μ 代表此课程的所有已评分用户的平均得分, b_u 是用户的偏置, b_i 是 item 的偏置。其中

$$b_i = \frac{\sum_{u \in R(i)} (r_{ui} - \mu - b_u)}{\lambda + |R(i)|} \quad (2)$$

$$b_u = \frac{\sum_{i \in R(u)} (r_{ui} - \mu - b_i)}{\lambda + |R(u)|} \quad (3)$$

6. 课程推荐系统的实现

6.1. 系统开发思想

在此课程推荐系统中, 每当用户对软件进行操作, 都将被用户日志信息服务器记录并反馈给数据库。由图 1 所示, 我们将用户的操作分成显式和隐式, 分别加以权重, 最后的用户评分将由这些合成。对所

记录的数据, 系统首先对数据进行去噪处理, 然后进行矩阵补全, 生成特征向量, 进而利用 KNNBaseline 实现离线推荐。

6.2. 算法步骤

步骤 1: 根据用户行为信息, 参考用户行为评级列表, 见表 3, 生成评分矩阵 P_D , 获取目标用户 u 的已评分矩阵 P_{DRated} 和目标用户 u 的未评分矩阵 $P_{\widehat{DRated}}$ 。

Table 3. The information of user rating table

表 3. 用户评级列表

UserBehavior	Type	DataType	Value	Description
Rating	Explicit	Integer	[0, 5]	The score people grade a course
Share	Explicit	Boolean	True or False	The status of whether people share a course
Download	Explicit	Boolean	True or False	The status of whether people download a course
Collect	Explicit	Boolean	True or False	The status of whether people collect a course
Comment	Explicit	RTF		The comment of people
Signup	Explicit	Boolean	True or False	The status of whether people sign up a course
Videorunningtime	Implicit	Double	[0, 10]	The time of people spend on a course

步骤 2: 分别进行目标用户 u 与同一学院 v 作和选了同一门课的同学作皮尔逊相似度计算, 使用基线 (而不是平均值) 计算所有对用户 (或项目) 之间的皮尔逊相关系数 (收缩), 收缩参数有助于避免在只有少量额定值可用时过度拟合。

$$sim(u, v) = \rho_{uv} = \frac{\sum_{i \in I_{uv}} (r_{ui} - b_{ui})(r_{vi} - b_{vi})}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - b_{ui})^2} \sqrt{\sum_{i \in I_{uv}} (r_{vi} - b_{vi})^2}} \quad (4)$$

$$sim(u, n) = \rho_{un} = \frac{\sum_{i \in I_{un}} (r_{ui} - b_{ui})(r_{ni} - b_{ni})}{\sqrt{\sum_{i \in U_{ij}} (r_{ui} - b_{ui})^2} \sqrt{\sum_{i \in U_{in}} (r_{ni} - b_{ni})^2}} \quad (5)$$

步骤 3: 目标用户与同学院的同学 KNNBaseline 计算, 得到推荐列表 L1,

$$r_{\widehat{ui}} = b_{ui} + \frac{\sum_{v \in N_i^k(u)} sim(u, v)(r_{vi} - b_{vi})}{\sum_{v \in N_i^k(u)} sim(u, v)} \quad (6)$$

步骤 4: 目标用户与选了同一门课的同学进行公式 2 计算, 得到推荐列表 L2,

$$r_{\widehat{ui}} = b_{ui} + \frac{\sum_{j \in N_u^k(u)} sim(u, n)(r_{uj} - b_{uj})}{\sum_{j \in N_u^k(u)} sim(u, n)} \quad (7)$$

步骤 5: 将推荐结果 L1 与 L2 取并集按分值从高到低排列, 根据已评分矩阵 P_{DRated} , 排除已评分课程, 将分值较高的 Top-4 作为最终的推荐结果。

7. 实验设计与结果分析

7.1. 实验准备

实验使用来自 GroupLens 的 MovieLens ml-100k 数据集, 数据集含有来自 942 名用户对 1682 部电影

的 10,000 条评分组成。每个用户至少评分 20 部电影。用户和电影从 1 号开始连续编号。数据是随机排序的, 这些数据分成四个列分别代表用户信息, 电影信息, 评分和时间。这与该系统的评分要求相类似, 电影信息包含了电影的类别, 这与课程的类别特征相对应。因此, 采集该数据集作为本系统的测试数据。[11]。

7.2. 实验结果分析

本实验将用十折交叉验证来验证该数据集的误差。十折交叉验证用来测试算法准确性, 是常用的测试方法。将数据集分成十份, 其中九份作为训练数据, 1 份作为测试数据, 进行实验。实验结果如图 7 所示。

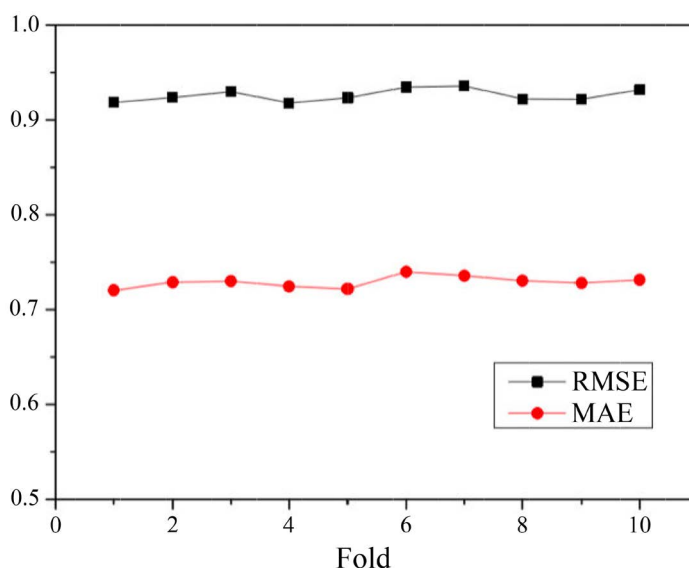


Figure 7. The result of experiment
图 7. 实验结果

7.3. 结果分析

该系统中, 推荐结果的准确率采用用户的预测评分值和真实评分值的平均绝对误差和均方根误差来评估, 平均绝对误差公式(7)和均方根误差公式(8)如下所示

$$MAE = \frac{1}{|R^{\wedge}|} \sum_{r_{ui} \in R^{\wedge}} |r_{ui} - r_{ui}^{\wedge}| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{|R^{\wedge}|} \sum_{r_{ui} \in R^{\wedge}} (r_{ui} - r_{ui}^{\wedge})^2} \quad (9)$$

分析其中 r_{ui} 表示用户对课程的预测评分, r_{ui} 代表用户对课程的实际评分。进行十折交叉验证。误差如图 1 所示。

由图可知, 该算法能良好预测用户对课程喜好度, 并能为该课程推荐系统产生不错的个性化推荐结果。

8. 结语

综上所述, 本文提出的基于协同过滤算法的移动智能学习平台, 目的是满足现代化高校在教学环节

中学生对个性化受教育的需求, 提高高校数字化教学水平, 实现了高校学生的个性化教育, 系统地整合了网课平台和教学管理辅助系统, 解决了当前市场教育类微信小程序分类复杂, 操作繁琐等特点, 具有界面风格清晰大方, 用户体验效果良好, 操作简单易懂, 操作实施方便的功能特点。该系统使高校教育和管理工作更加科学化、规范化, 在现实中具有良好的应用前景。

参考文献

- [1] 极光大数据: 2017年8月在线教育市场研究报告[EB/OL]. <http://www.199it.com/archives/637393.html>, 2017-09-26.
- [2] 2018 微信用户数据报告[EB/OL]. <https://www.useit.com.cn/thread-20174-1-1.html>, 2018-08-27.
- [3] 汪志华, 范祥林, 郁书好. 高校实验室管理系统开发与设计[J]. 软件工程与应用, 2018, 7(4): 218-223.
- [4] 王渊博, 陈有波. 农产品直卖平台服务端软件设计[J]. 现代电子技术, 2019(8): 171-174.
- [5] 张洪伟. JSP 网络开发技术与整合应用[M]. 北京: 清华大学出版社, 2006.
- [6] 宋雨欣. 医院预约挂号平台的设计与实现[D]: [硕士学位论文]. 郑州: 郑州大学, 2016.
- [7] 赵鑫. 医院预约挂号平台的设计与实现论文[EB/OL]. <https://wenku.baidu.com/view/0928433ab94ae45c3b3567ec102de2bd9605dec4.html>, 2017-09-08.
- [8] 安忠, 佟志臣. 管理信息系统[M]. 第二版. 北京: 中国铁道出版社, 2006.
- [9] 刘莹. 基于微信的高校自主学习平台设计与实现[J]. 中国成人教育, 2017(20): 18-21.
- [10] Koren, Y. (2010) Factor in the Neighbors: Scalable and Accurate Collaborative Filtering. *ACM Transactions on Knowledge Discovery from Data*, 4, 1-24.
- [11] 徐林. 基于 Spark MLlib 协同过滤算法的美食推荐系统研究[J]. 吉林大学学报(信息科学版), 2019, 37(2): 181-185.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2286, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sea@hanspub.org