

DOI:10.16644/j.cnki.cn33-1094/tp.2021.07.012

面向个性化学习的慕课资源推荐系统开发*

孔令圆, 彭 琰, 郑汀华, 马 华

(湖南师范大学信息科学与工程学院, 湖南 长沙 410081)

摘 要: 近年来, 在线学习得到了大规模普及, 互联网上已发布了海量的慕课学习资源。针对广大学习者进行在线慕课学习时面临的“信息迷航”和“信息过载”等问题, 以混合推荐算法为基础, 通过多角度挖掘用户的个性化信息, 设计并实现了一个面向个性化学习的慕课资源推荐系统, 以此为广大学习者提供自主学习的辅助支持。该系统能充分挖掘学习者的显式和隐式偏好, 为其推荐满意的慕课资源, 具有良好的应用价值。

关键词: 慕课; 学习资源; 混合推荐; 个性化学习

中图分类号: G642

文献标识码: A

文章编号: 1006-8228(2021)07-46-04

Development of MOOC resources recommendation system for personalized learning

Kong Lingyuan, Peng Yan, Zheng Tinghua, Ma Hua

(College of Information Science and Engineering, Hunan Normal University, Changsha, Hunan 410081, China)

Abstract: In recent years, online learning has been popularized on a large scale, and massive MOOC learning resources have been released on the Internet. Aiming at the problems of "information disorientation" and "information overload" faced by the majority of learners in online MOOC learning, on the basis of hybrid recommendation algorithm, by mining user's personalized information from multiple perspectives, a MOOC resource recommendation system for personalized learning is designed and implemented, in order to provide the majority of learners with the auxiliary support of autonomous learning. The system can fully tap learners' explicit and implicit preferences, and recommend satisfactory MOOC resources for them, which have a good application value.

Key words: MOOC; learning resource; hybrid recommendation; personalized learning

0 引言

近年来, 国家的一系列指导性文件^[1-2]明确指出了以“学习者”为中心、个性化培养的重要性, 有力地推动了我国在线教育的发展。目前, 主流在线教育平台已在互联网上发布了海量的慕课学习资源, 在线学习者人数增长迅速。而2020年初暴发的新冠疫情进一步推动了在线学习的大规模普及。

面对海量的慕课学习资源, 广大学习者在慕课学习时面临着“信息迷航”和“信息过载”^[3-4]等问题, 学生很难充分了解并快速挑选出最适合自己的学习资源。由此, 实现学生需求与学习资源的匹配, 从而提高学生进行在线慕课学习时的学习效率和在线教育平台

中慕课学习资源的利用率, 具有重要的现实意义。

针对当前主流在线教育网站缺乏慕课资源推荐功能的不足, 本文在传统混合推荐算法的基础上, 多角度挖掘用户的个性化信息, 设计并开发了一个面向个性化学习的慕课资源推荐系统, 以此为广大学习者提供高效的自主学习的辅助支持。

1 混合推荐算法原理

基于内容的推荐算法通过抽取物品的内容信息来描述物品并形成物品的特征模型^[5], 培训尝试向用户推荐与其过去所喜好物品在内容上相似的其他物品。协同过滤算法基于兴趣相投的假设, 借助于用户

收稿日期: 2021-02-25

*基金项目: 湖南师范大学智能计算与感知类大学生创新创业训练项目(No. 1); 湖南省普通高等学校教学改革研究项目(湘教通(2019)291号, No. 140)

作者简介: 孔令圆(1999-), 女, 浙江金华人, 本科生, 主要研究方向: 个性化学习与推荐。

通讯作者: 马华(1979-), 男, 湖南株洲人, 博士, 副教授, 主要研究方向: 智慧教育和个性化学习。

群体的行为模式分析,以合作的方式进行项目过滤。该推荐算法主要是对用户群体的行为数据进行分析,寻找邻近用户或邻近物品进而生成推荐^[6],如公式(1)常用于基于用户的协同过滤算法的评分预测任务。将以上两种类型的推荐算法使用一定的机制相混合,便得到了基于用户反馈和物品内容的混合推荐算法^[7]。

$$\hat{r}_{u,j} = \bar{r}_u + \frac{\sum_{v \in N} \text{sim}(u,v) \times (r_{v,j} - \bar{r}_v)}{|\sum_{v \in N} \text{sim}(u,v)|} \quad (1)$$

其中, $\hat{r}_{u,j}$ 表示用户 u 对物品 j 的预测评分; $r_{v,j}$ 是用户 v 对物品 j 的实际评分; \bar{r}_u 为用户 u 的平均评分; N 是评价过物品 j 的相似用户群; $\text{sim}(u,v)$ 为用户 u 和 v 的相似度。

2 个性化慕课推荐算法流程

2.1 基于标签的课程建模

将经过预处理后的课程标签数据,转换为 Python 中的字典格式,得到课程模型组。转换得到课程模型组样式如下:

```
courses_models = { courseid1: {'label1':1,'label2':
0.8,...}, courseid2: {...} ... }
```

2.2 基于标签的用户建模

结合用户课程评分数据和课程建模得到的课程模型组,由公式(2)计算用户对标签的评分(即喜爱程度)。

$$R_{u,i} = \text{Sig}(\sum_{j \in N} \text{Sim}_{ij}) \quad (2)$$

其中, $R_{u,i}$ 表示用户 u 对标签 i 的评分; Sig 表示 sigmoid 变体函数; N 表示用户 u 已看且与标签 i 相关的课程集合; Sim_{ij} 表示标签 i 与课程 j 的相关系数。进而得到用户模型组,其样式如下:

```
users_models = { user1: {'label1':0.8,'label2':0.5},
user2: {...} ... }
```

公式(2)针对用户 u 已看且与标签 i 相关的所有课程,将这些课程与标签 i 的相关系数求和,再使用 sigmoid 变体函数进行归一化处理,得到用户 u 对标签 i 的评分。这里没有将用户 u 对已看课程 j 的实际评分作为对标签 i 评分的影响因素,是因为用户对具体某一门课程的评分往往涉及许多除开课程内容之外的其他因素,如教师授课质量,课程计划安排等。相比之下,只考虑用户对某一标签相关课程的访问频次信息,能更准确反映用户对该标签的喜爱或需求程度。sigmoid 变体函数如公式(3)所示。

$$\text{Sig}(x) = \frac{1}{1 + e^{-2x + 4.8}} \quad (3)$$

2.3 基于综合相似度的混合推荐

首先,计算基于标签的用户模型与课程模型之间的余弦相似度,挑选相似度较高的前 N 门课程作为初始推荐列表。然后,将前 N 门课程与用户的相似度作为其初始推荐系数。计算当前用户与其他所有用户的“综合相似度”后,设置合理阈值,得到相似用户组。使用相似用户的课程评分数据,调整当前用户推荐列表中各课程的推荐系数。其中,“综合相似度”由评分反馈相似度、专业相似度、年级相似度和兴趣相似度加权组合而成。

2.4 结合注册信息的推荐

对于标题或简介里含有当前用户兴趣关键词的课程,增加一定比例的推荐系数。同时,结合当前用户与课程的层次匹配度调整各门课程的推荐系数。层次匹配度的计算着重参考了全国第四轮学科评估的数据(https://souky.eol.cn/api/newapi/assess_result)和 2020 年校友会中国大学排名数据(http://report.iresearch.cn/report_pdf.aspx?id=3728.2020 April)。将学科评估为 A+ 类的院校排名设为 1, A 类院校排名设为 2, 依此类推。当用户所在专业与课程所属学科一致且用户所在院校与课程开办院校均有对应的学科评估排名时,我们按照学科评估排名的前后位序差距进行等距计算并赋予相应权重,使得两者学科评估排名越相近,则层次匹配度越高。除上述情况外,我们按照院校综合排名的前后位序差距进行等距计算并赋予合理权重,使得两者院校综合排名越相近,则层次匹配度越高。

2.5 结合浏览痕迹的推荐

首先,结合搜索痕迹来影响课程的推荐系数。对于标题或简介里含有搜索关键词的课程,增加一定比例的推荐系数。其次,结合停留痕迹信息来影响课程的推荐系数。为与停留页面信息相关的课程即标题或简介中含有停留页面关键词的课程,提供与对应停留时间呈线性相关的推荐系数提升比例。

3 慕课资源推荐系统的开发

3.1 基于 Hibernate 和 MVC 的总体设计

3.1.1 基于 Hibernate 框架的持久化设计

(1) 建立持久化课程类和学生类。课程类包括 id、名称、开办院校、授课教师、选课人数、简介、链接等属

性;学生类包括id、姓名、密码、性别、所在院校、年级、所在专业、兴趣爱好等属性。

(2) 建立持久化类映射文件(Course.hbm.xml、Student.hbm.xml),将课程类、学生类分别与数据库中的课程详细信息表、学生信息表关联起来。

(3) 建立Hibernate配置文件,配置连接数据库的参数和持久化类映射文件的位置。

(4) 建立Hibernate工具类,封装对会话工厂对象和会话对象的创建和维护。

(5) 在控制器servlet中对数据表进行查询,得到持久化对象列表。

3.1.2 基于MVC模式的系统框架设计

(1) 模型为课程类、学生类。

(2) 视图包括综合推荐页(select-class.jsp)、课程

搜索页(search-class.jsp)、个人中心页(individual.jsp),它们负责向用户呈现JavaBeans中的数据。

(3) 控制器为mainervlet以及它涉及的7个业务处理类。业务处理类包括综合推荐处理类、知识元推荐处理类、推荐列表检索处理类、慕课搜索处理类、登录处理类、注册处理类、个人中心访问处理类。

3.2 系统组成与主要功能模块

根据上述设计,我们实现了一个真实的个性化慕课推荐系统(<http://115.159.54.233:8080/moocrec/index.jsp>),并已将其公开部署于腾讯云服务器。整个系统共包含热门推荐、综合推荐、知识元推荐、慕课搜索、登录、注册和个人中心等七个模块。系统的功能模块结构如图1所示。该系统的首页运行界面及推荐结果界面分别如图2和图3所示。

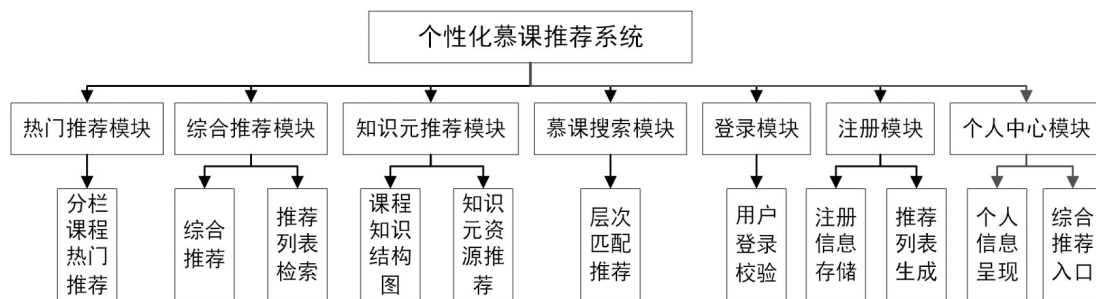


图1 系统的功能模块结构



图2 系统首页

课程名称	课程简介	开办院校	教师	热度
大学计算机: Python语言实践	本课程是一个入门课程, 旨在为初学者提供Python语言的基本知识。	东南大学	马勇	11368
Visual Basic.NET程序设计入门	5000名学员好评, 涵盖了从基础到高级的知识点。	同济大学	陈国良	3873
Visual Basic.NET程序设计进阶	5000名学员好评, 涵盖了从基础到高级的知识点。	同济大学	李江全	5119
最新流行程序设计: Python	“最新流行程序设计”课程旨在帮助初学者快速掌握Python语言。	哈尔滨工业大学	陈海源	70641
最新流行程序设计: Python CAP	“最新流行程序设计”课程旨在帮助初学者快速掌握Python语言。	哈尔滨工业大学	宋海峰	79587
最新流行程序设计: Python	本课程旨在帮助初学者快速掌握Python语言。	哈尔滨工业大学	陈海源	17528
Java核心技术	Java是一门跨平台设计语言, 长期受到全球开发者的广泛使用。	华东师范大学	王永生	10148
Visual Basic.NET程序设计	本课程旨在帮助初学者快速掌握Visual Basic.NET语言。	同济大学	李江全	11054
Python语言程序设计	计算机专业工具, 课程旨在帮助初学者快速掌握Python语言。	北京理工大学	戴天	134765

图3 系统推荐结果页

七个功能模块的设计和实现特点分别介绍如下。

3.2.1 热门推荐模块

将爬取的课程数据划分为十二个大类, 分别为计算机、物理与电子、化学与化工、生物医学、历史与文学、数学与统计、经济、管理、教育与心理学、外语、艺术及政法等类别。在首页的热门推荐板块, 利用热门课程信息和课程链接为用户提供热门课程入口。

3.2.2 综合推荐模块

当用户请求获得综合推荐时, 该模块将访问数据库的推荐列表缓存表, 得到推荐课程的id列表。结合Hibernate框架将推荐列表由课程id列表转化为课程持久化对象列表后, 在前端综合推荐页向用户呈现推荐结果。通过定期调用更新模块, 数据库中的推荐列表缓存表将不断更新, 充分结合用户新的行为痕迹信息影响综合推荐, 以适应用户的新偏好。

3.2.3 知识元推荐模块

我们使用百度脑图工具制作了内含学习资源链接的知识结构图, 这些图存储在百度云端的海量空间上。通过生成共享链接, 可将知识结构图嵌入本系统

的课程详情页,以实现知识元资源推荐的可视化展示。针对每门课程,我们在百度云端存储了三种不同难度层次的知识结构图,它们关联了对应学科大类下 A、B、C 类三类不同院校开办的课程,并以它们来匹配处于不同层次的 A、B、C 三类用户。A 类和 B 类用户分别对应 A 类和 B 类院校的学生, C 类用户对应 C 类院校或未参评院校的学生。从而实现为同一专业但不同学科层次的学生提供不同难度的知识元资源索引服务。以“数据结构”课程为例,层次难度 1 的知识结构图挂接了北京大学和浙江大学的课程大纲,而各个知识点可链接至这两所大学发布的学习资源;系统会自动为所学专业(计算机类)评估为 A 类的学生呈现此知识结构图。

此模块利用了前端 AJAX 技术和后端 jFinal 框架。用户刷新或关闭课程的详情页时,前端 JavaScript 脚本将页面 URL、停留时间、用户开始访问时间和结束访问时间、用户 id 保存为 JSON 字符串,并利用 AJAX 向后端传递此数据。后端控制器接收并解析 JSON 字符串,将数据存入对应模型并保存至数据库。

3.2.4 慕课搜索模块

该模块用于提供对海量数据的语义检索功能。用户可在慕课搜索页的搜索框中输入关键词,系统将向用户提供与关键词语义相关的课程。系统会结合院校排名数据,根据排名的前后位序差距进行等距计算,优先为用户呈现层次需求匹配的课程。同时,此模块记录了用户的搜索词信息,用于更新综合推荐。

3.2.5 登录模块

用户提交登录信息后,系统将结合 Hibernate 框架访问数据库获取用户持久化对象。当该持久化对象不为空且用户名与密码校验正确时,用户才能登录系统,以便系统根据其个性化信息提供综合推荐功能、知识元推荐功能和慕课搜索功能。

3.2.6 注册模块

用户提交注册信息后,系统将结合 Hibernate 框架为用户生成持久化对象并存入数据库。随后系统启用多线程机制,主线程负责用户后续登录服务,子线程负责调用推荐算法为新用户生成综合推荐结果并存入数据库。

3.2.7 个人中心模块

该模块向用户呈现其详细的个人信息,包括用户名、在读院校、所在年级、在读专业、兴趣爱好等,并向用户提供综合推荐访问入口。该模块呈现的信息,为系统综合推荐、知识元推荐和慕课搜索的结果提供一定的可解释性。

4 结束语

本文开发的面向个性化学习的慕课资源推荐系统,是以混合推荐算法为基础,使用了 Hibernate 和 MVCWeb 开发技术,可动态捕获学习者的在线行为数据,能从多角度感知用户特征,是一个操作简单、可为学习者提供高效支持的个性化辅助学习工具。该系统充分挖掘了用户的偏好与潜在需求,能较好地解决学习者在线学习时面临的信息迷航和信息过载问题,具有良好的应用价值。未来研究中,我们拟引入知识图谱技术,对慕课资源所属的知识体系及知识点间的关联关系建模,从而为学习者提供可视化的学习路径规划服务和具有更高满意度的推荐结果。

参考文献(References):

- [1] 中华人民共和国教育部. 教育部关于印发《教育信息化 2.0 行动计划》的通知[EB/OL]. http://www.moe.gov.cn/srcsite/A16/s3342/201804/t20180425_334188.html
- [2] 新华网. 中共中央、国务院印发《中国教育现代化 2035》[EB/OL]. http://www.moe.gov.cn/jyb_xwfb/s6052/moe_838/201902/t20190223_370857.html
- [3] Shanshan Wan, Zhendong Niu. An e-learning recommendation approach based on the self-organization of learning resource[J]. Knowledge Based Systems, 2018. 160: 71-87
- [4] Robert Bodily, Katrien Verbert. Review of research on student-facing learning analytics dashboards and educational recommender systems[J]. IEEE Transactions on Learning Technologies, 2017. 10(4): 405-418
- [5] 刘宇,朱文浩. 基于内容和标签权重的混合推荐算法[J]. 计算机与数字工程, 2020. 48(4): 773-777
- [6] 刘佳奇,王全民. 基于改进的用户协同过滤算法的高校个性化图书推荐系统[J]. 计算机与数字工程, 2020. 48(10): 2458-2461, 2479
- [7] 郭贵冰. 推荐系统进展: 方法与技术[M]. 科学出版社, 2018. 