

论开放教育教学资源的智能推荐技术

刘海涛,段小芳,戴苏南
(江苏开放大学 南通学院,江苏 南通 226000)

摘要:教学资源是现代教育教学的重要素材,随着教学资源越来越丰富,学生如何快速查找或发现感兴趣的教学资源就显得越来越重要。在构建教学资源模型和用户模型的基础上,针对教学资源和用户特点,采用基于用户和项目推荐算法的混合协同智能推荐算法来实现教学资源的智能推荐。

关键词:教学资源;资源相似度;混合协同推荐算法
中图分类号:G724.82 **文献标识码:**A **文章编号:**1008-5092(2017)04-0081-04

网络和多媒体技术的发展掀起了教育信息化热潮,网络教学资源的建设及使用,极大影响了各类教学的组织形式、教学方法和教育理念。在此背景下,建立教学资源管理平台,可以方便学生通过网络来获取学习资源以及参与教学讨论。但“资源过载”和“资源迷航”现象使得学生搜索所需资源的成本越来越高^[1]。为解决学生群体特殊性、学生关心的教学资源的差异性问题,迫切需要使用智能推荐技术,来实现教学资源的智能推荐,帮助学生挖掘潜在的学习兴趣,让用户能快速发现有价值的资源,提高知识推送的效率。

一、教学资源与用户模型构建

1. 教学资源模型构建

教学资源模型用于描述数据库中的各类在教学中使用的资源,教学资源模型表如表 1 所示。

2. 用户与模型构建

在教学资源平台中,其用户群体主要包括教师和学生两类。分析学历教育学生或非学历教育用户的学习背景可以知道,这些学习者都具备学历层次或专业(行业)背景,因此教学资源平台中的用户群体要较基于 Web 的推荐系统中的用户群体更加稳定。

根据教育教学的特点,教学资源平台推荐算

法的学生信息描述如表 2 所示。

表 1 教学资源模型表 Table 1 Teaching resources model	
属性	描述
资源编号	主键,教学资源的 ID
资源类型	分为:网络课程知识、典型问题解答、案例、测试题库、媒体素材、文献资料、资源目录索引 ^[2]
专业代码	参照《中华人民共和国国家标准学科分类与代码(GB/T13745-2009)》
资源名称	教学资源的名称描述
教育类型	分为:全日制学历教育(子分类:专、本、硕、博);远程开放学历教育(专科、本科);社会教育
内容简介	教学资源的简单文字描述
作者	教学资源的作者
上传时间	教学资源上传到资源库的时间

表 2 学生信息描述表 Table 1 Students' information description	
属性	描述
学号	主键,学生 ID
专业代码	参照《中华人民共和国国家标准学科分类与代码(GB/T13745-2009)》
教育类型	同表 1 中的教育类型分类
学生姓名	学生的姓名
性别	学生的性别描述

在学生用户模型和教学资源模型中,通过专业代码,使学生用户和教学资源之间构成某种特定的关联关系。但是,在学生学习过程中,学生除了对本专业的教学资源感兴趣以外,还有可能有其它的兴趣和爱好,即学生跨界学习时可能会关注除了其本专业以外的其它教学资源。

在教学资源推荐算法中,使用用户-项目矩阵来描述用户对平台中教学资源感兴趣的程度。假设教学资源管理平台中的用户群体集合表示为 $U = \{U_1, U_2, \dots, U_m\}$, 平台中的教学资源项目集合表示为: $D = \{D_1, d_2, \dots, D_n\}$ 。在打分矩阵 $R =$

$\{r_{ij}\}_{n \times m}$ 中的元素 r_{ij} 表示用户 U_j 对教学资源 D_i 的感兴趣程度^[2]。

二、教学资源推荐流程

教学资源平台中的教学资源种类繁多,可以通过教学资源模型中的“专业代码”获得教学资源分类,结合目标用户的“专业代码”属性,将教学资源分为用户本专业教学资源和非本专业教学资源两类。同时又在此基础上分为最新的资源、用户感兴趣的资源和精品资源等类别,其具体的分类结构图如图 1 所示。

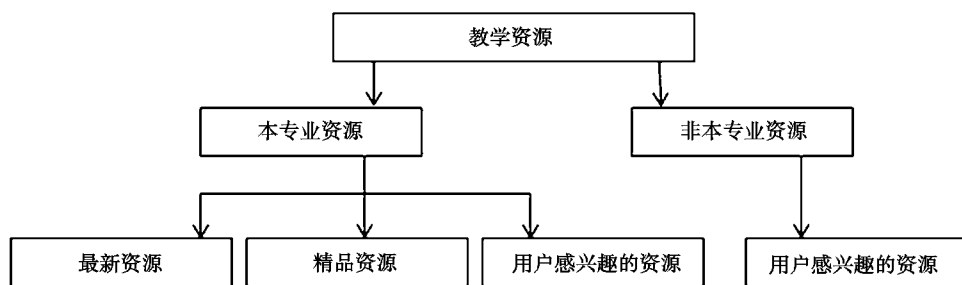


图 1 教学资源分类图

Fig. 1 Teaching resources classification

1. 本专业资源推荐

(1) 精品教学资源的推荐

如果一个教学资源获得了足够多的评价之后,就可以通过用户对该教学资源的评价情况来衡量该教学资源的质量,即如果平均值较高,则表明该资源质量更好,预测其它用户可能也会对该教学资源感兴趣,操作时将平台中与用户专业匹配、用户评分值较高的教学资源作为推荐项推荐给客户。

(2) 最新资源的推荐

在教学资源管理过程中,教学资源的制作者会源源不断地上传教学资源,由于新的教学资源还未有评分用户,或者评分用户较少,因此导致一些质量不错但上传时间短的教学资源,在较长的一段时间内难以得到推荐。为此,在智能推荐算法中应该考虑将最新上传的教学资源也推荐给客户。

(3) 用户感兴趣的教学资源

在前面两种教学资源的推荐过程中,都是从教学资源本身的属性来判断是否需要推荐的。而用户感兴趣的教学资源的推荐,就需要通过对用户相似性和教学资源项目相似性的计算,使用协同过滤算法,来判断是否进行教学资源的

推荐。

2. 非本专业教学资源推荐

对于学生用户而言,他们关心的重点是本专业的教学资源。但事实上,学生为丰富综合知识,往往会关注其他相近专业或风牛马不相及的专业课程,希望通过提供的教学资源跨界了解一些其它知识。为此,在选择、改进智能推荐算法时,首先考虑让用户进入教学资源管理平台注册模块时选择兴趣爱好;平台系统收集用户可能感兴趣的专业。然后通过运用挖掘方法分析用户使用资源平台的历史数据,预测用户的兴趣点,从而有效实现向用户推荐匹配的教学资源。基于用户的协同过滤的推荐结果着重于反映和用户兴趣相似的小群体的热点,而基于项目的协同过滤的推荐结果着重于维系用户的历史兴趣^[3]。由于非专业教学资源推荐的资源分类不确定,因此在进行非专业教学资源的推荐时,既需要通过用户已有的打分信息来发掘用户潜在的兴趣爱好,也需要考虑教学资源项目的相似性,结合基于用户推荐算法和基于项目推荐算法的优点,采用混合协同过滤算法来进行教学资源的推荐。非本专业的教学资源推荐流程如图 2 所示。

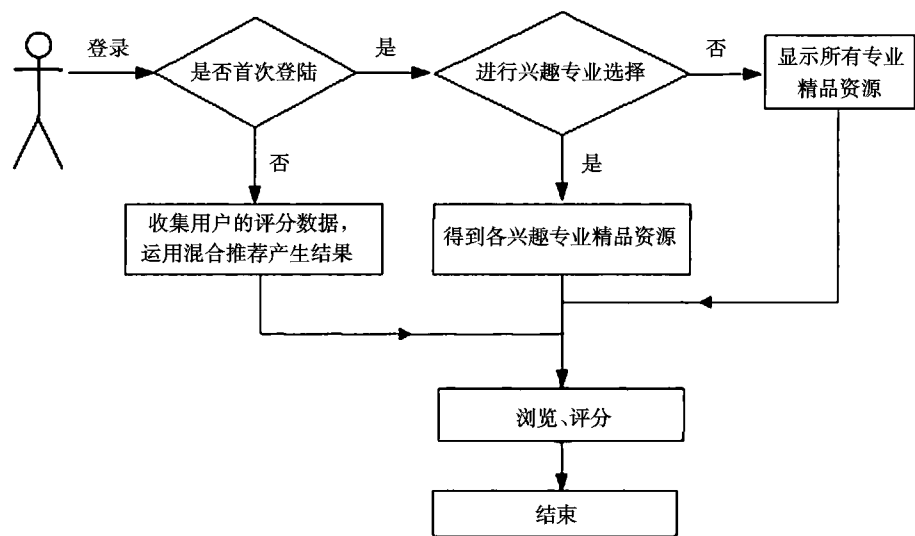


图 2 非本专业教学资源推荐流程图

Fig. 2 Recommended flowchart of teaching resources for non – professional teaching

三、相似度的计算算法分析

1. 相似度计算存在的难点

目前应用较广的相似度计算方法主要有 Pearson 相似度计算算法和余弦相似度算法。这些应用较为广泛的相似度计算方法已经非常成熟,但是应用到教学资源智能推荐算法中时仍然存在一些不足。其主要表现为:随着教学资源数量的不断增多,系统中用户数量迅速增长,以及平台中用户参与度提高,会导致用户对教学资源的评价矩阵越来越稀疏,即用户仅对平台中较少的教学资源有评价,而其它绝大部分的教学资源不进行评价。借鉴大型电子商务系统的经验,用户评分项不会超过总项目的 1%^[4],而两个用户共同的教学资源评分项则会更小。为此,在教学资源平台的智能推荐算法中,用户评分矩阵可能存在极端稀疏的情况,这时再使用传统的 Pearson 相似度计算算法和余弦相似度算法就不再合适。

在使用传统的相似度计算方法进行用户相似度计算时,如果用户没有对教学资源项目进行评价,则在用户 – 项目评分矩阵中的相应元素取值为 0。假设用户的评价的项目项不超过 1%,那么两个用户各有 99% 的项目评价结果都为 0,即这两个用户最少有 98% 的相似度,这显然不能说明两个用户的相似程度。用户相似度计算方法的不准确,导致邻居用户的选择不合理,导致课程智能推荐质量急剧下降,影响知识推送的效果。

2. 用户相似度计算算法设计

针对上述的教学资源平台智能推荐算法中的

用户相似度计算不准确、邻居用户选择不合理的问题,本文提出一种改进的相似度计算方法。例如,在教学资源项目的两两相似度计算时,首先计算对两个项目进行共同评分的用户集合并集 $U_{ij} = U_i \cup U_j$,其中, U_i 表示对教学资源项目 D_i 进行评分的所有用户集合, U_j 表示对教学资源项目 D_j 进行评分的所有用户集合。在用户集合 U_{ij} 中,没有对教学资源项目 D_i 或教学资源项目 D_j 进行评分的用户,可以通过与其它用户的相似性预测其评分,然后计算教学资源项目 D_i 和教学资源项目 D_j 的相似度。这种相似度计算方法有效解决了教学资源智能推荐算法中的稀疏性问题,同时也可以解决采用 Pearson 相似性计算方法和余弦相似性计算方法因未评分项元素值相同导致的推荐质量下降问题。

四、教学资源混合协同过滤推荐算法设计

1. 获取用户 – 项目评分矩阵

用户使用教学资源后,会对资源进行评价。其评价信息以矩阵 $R = \{r_{ij}\}_{n \times m}$ 的形式表示,其中 m 列表示了系统中的 m 个用户, n 行表示了系统中的 n 个教学资源项目。矩阵中的元素 r_{ij} 表示用户 U_j 对教学资源 D_i 的评分结果,即感兴趣程度或满意程度。

2. 相似项目集形成

对于系统中的任意两个教学资源项目 D_i 和 D_j ,使用如前所示的改进的相似度计算方法计算

出两个教学资源项目之间的相似度 $sim(i, j)$ 。通过相似度两两之间的计算,构建教学资源项目 D_i 与其最相似的 N 个教学资源项目组成集合 SI_i 。

3. 邻居用户的寻找

寻找用户 U_u 与用户 U_j 分别进行评分评价的教学资源集合 $I_{ij} = I_i \cup I_j$, 而对于未评分项, 则通过未评分教学资源项目与用于已评分教学资源项目之间的相似度的计算, 以及预测用户对未评分教学资源项目的评分结果^[5], 不断循环上述的过程, 直到所有的预测值都填充到评分矩阵中。

在完成用户-项目评价矩阵的填充之后, 就可以使用 *Pearson* 相似度计算方法来计算用户 U_i 与用户 U_j 之间的相似度 $sim(i, j)$ 。

$$sim(i, j) = \frac{\sum_{k \in I_{ij}} (R_{i,k} - \bar{R}_i)(R_{j,k} - \bar{R}_j)}{\sqrt{\sum_{k \in I_{ij}} (R_{i,k} - \bar{R}_i)^2} \sqrt{\sum_{k \in I_{ij}} (R_{j,k} - \bar{R}_j)^2}}$$

通过计算用户之间的相似度, 构建用户 U_i 与

其最相似的 N 个 *yonghu* 组成集合 N_i , $\bar{R}_i$ 表示用户 i 对资源的平均得分。

4. 产生推荐

计算用户 U_i 对未评测教学资源项目 D_j 的预测评分计算为:

$$P(i, j) = \bar{R}_i \frac{\sum_{n \in N_i} sim(i, n) \cdot (R_{n,j} - \bar{R}_n)}{\sum_{n \in N_i} sim(i, n)}$$

通过对用户 U_i 对所有未评分项目的预测评分, 选择预测与用户最相近的 N 个教学资源项目, 形成 *top-N* 推荐集推荐给用户 U_i 。

本文主要在信息化教学快速发展及广泛运用的背景下, 为学习者解决所需信息与搜索之间的矛盾^[6], 在分析了常见智能推荐算法的基础上, 设计了一种基于用户和基于项目的混合推荐算法来实现教学资源的智能推荐, 以更好地提高教学资源管理平台的用户体验和教学资源管理平台中的资源利用率。

参考文献:

- [1] 黎孟雄, 郭鹏飞. 基于模糊聚类的教学资源自适应推荐研究[J]. 中国远程教育, 2012, 7(5): 89-90.
- [2] 李俊薇. 基于协同过滤的校园教育资源网个性化推荐研究[D]. 武汉: 华中师范大学, 2009.
- [3] 蒋漪涟, 刘晓丹, 张路通. 教学资源推荐系统的研究与设计[J]. 微型电脑应用, 2016, 32(6): 20-23.
- [4] 乌兰, 于洪涛. 教育技术学专业教材建设存在的问题及建议[J]. 民族高等教育研究, 2016, 4(1): 77-82.
- [5] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1624-1626.
- [6] 李玮, 何富乐, 骆嘉伟, 等. 支撑教学资源的推荐算法原型系统设计[J]. 中国教育信息化, 2016(5): 33-35.

Research on Intelligent Recommendation Technology of Teaching Resources

LIU Haitao, DUAN Xiaofang, DAI Sunan

(Nantong College, Jiangsu Open University, Nantong Jiangsu 226000, China)

Abstract: With the deepening of information technology, online teaching resources are becoming more and more abundant. It is increasingly important to find the most suitable learning resources in "Resource Ocean". According to the openness of open educational resources, based on the construction of teaching resources model and users' model, following the basic characteristics of teaching resources and users, the paper adopts hybrid collaborative recommendation algorithm based on recommendation algorithm of users and projects so as to realize the intelligent recommendation of teaching resources. It has certain practical significance to solve the problems such as group specificity and the diversity of students' concern on teaching resources, tap students' potential interest in learning and enable users to find valuable resources quickly so as to improve the efficiency of knowledge push.

Keywords: teaching resources; resource similarity; hybrid collaborative recommendation algorithm

(责任编辑: 洪 林)