

# 基于协同过滤的个性化推荐选课系统研究\*

徐天伟<sup>1,2</sup> 宋雅婷<sup>2</sup> 段崇江<sup>2</sup>

(1. 华中科技大学 教育科学研究院, 湖北武汉 430074; 2. 云南师范大学 信息学院, 云南昆明 650500)

**摘要:**针对高等学校学生选课系统中存在的缺乏个性化课程推荐、选课效率不高的问题,通过对个性化推荐技术的分析研究,提出了基于项目、用户及属性值矩阵的协同过滤算法,并把该算法应用到选课系统中,数据表明,算法解决了个性化推荐技术中的冷启动问题,相关指标有明显提高,实现了课程的个性化推荐和新课程的推荐。

**关键词:**协同过滤;个性化推荐;选课系统

【中图分类号】G40-057 【文献标识码】A 【论文编号】1009—8097(2014)06—0092—07 【DOI】10.3969/j.issn.1009-8097.2014.06.014

## 引言

互联网的迅速发展带来了网上信息量快速、大幅增长,使得一般用户在海量信息中无法快速、准确获得所需信息,对信息的利用率反而降低,信息超载的现象广泛存在,解决办法之一是以搜索引擎为代表的信息检索系统,比如 Google、Baidu 等,但其基于关键字的搜索得到的结果不理想;另一有效、有潜力的办法是个性化推荐系统,它是根据用户的信息需求、兴趣等,将用户感兴趣的信息、产品等推荐给用户的个性化信息推荐系统。与搜索引擎相比,推荐系统通过研究用户数据,进行个性化计算,发现用户兴趣偏好,从而引导用户发现自己的信息需求<sup>[1]</sup>。个性化推荐系统是根据用户的兴趣爱好,推荐符合用户个性的对象<sup>[2]</sup>。个性化推荐系统已广泛应用于各个领域,斯坦福大学的个性化导航系统 Web-Watcher 是个性化信息服务的标志<sup>[3]</sup>,最典型并具有良好的发展和应用前景的是电子商务领域。

在高校的教学管理实践中为学生提供了大量的选修课程,门数众多,并且存在课程门类、专业等结构性的不足和缺憾,在诸如课程资源中心的组织和管理方式及现今多数的选课方式下,学生难以选择到适合的、符合个人专业发展及个性需求的课程。目前国内不少高校实际实施的是不完全学分制,由于资源的限制一半以上的课程都是必修课<sup>[4]</sup>,学生选课的余地较小,而且大多数的选课系统提供的仅仅是选课功能,缺乏为学生推荐课程的个性化选课功能。鉴于此,我们将个性化推荐技术应用到选课系统中,根据学生的学习需求和兴趣偏好,为学生提供个性化的选课推荐,从而避免学生选课的盲目性,提高课程资源的利用率和选课质量。

目前,个性化推荐技术还存在诸多问题,如数据稀疏问题,冷启动问题<sup>[1]</sup>。本文针对协同过滤推荐的数据稀疏性和冷启动问题,提出了基于项目、用户及属性值矩阵的协同过滤推荐算法来给用户推荐个性化的课程,满足学生学习的个性化需求,这样不仅可以提高课程资源的利用率,还能调动学生的学习兴趣和积极性。

## 一 协同过滤技术

协同过滤推荐(Collaborative Filtering Recommendation)技术是个性化推荐技术中最为成功的技术之一,于1992年由Goldberg等学者在研究报告中正式提出。1996年,美国明尼苏达大学(the University of Minnesota)的研究人员构建的MovieLens推荐系统就是利用协同过滤的原理来

推荐电影，并在当年发布到网上，可以说这是协同过滤领域的另一个开创性的工作<sup>[5]</sup>。典型的系统有 Jester 笑话推荐系统<sup>[6]</sup>、Ringo 音乐推荐系统<sup>[7]</sup>、GroupLens 网上新闻过滤系统<sup>[8]</sup>、亚马逊网上书店等。

协同过滤技术的基本思想是：利用一些兴趣爱好相似、拥有相同经验群体的偏好给目标用户推荐可能感兴趣的项目。协同过滤算法的实现大致分为构建用户——项目矩阵，寻找最近邻居，产生 TOP-N 推荐数据三个步骤<sup>[9]</sup>。目前协同过滤推荐技术的主要类型为：基于用户的协同过滤推荐、基于项目的协同过滤推荐和基于模型的协同过滤推荐。

### 1 基于用户的协同过滤推荐方法<sup>[10]</sup>

基于用户的协同过滤推荐算法是根据用户访问行为的相似性向目标用户推荐可能感兴趣的资源，其处理过程如图所示。

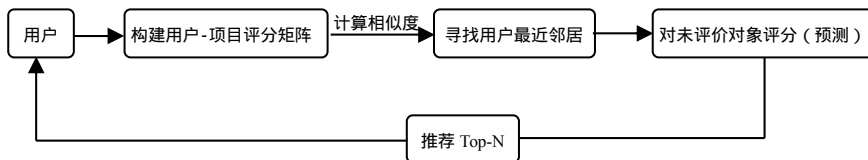


图1 基于用户的协同过滤推荐过程流程图

目前用户间相似性的度量方法主要采用修正余弦相似性函数。假设计算用户 a 和 b 之间的相似性，如公式 3<sup>[11]</sup>。

$$\text{Sim}(a, b) = \frac{\sum_{c \in I_{a,b}} (R_{a,c} - \bar{R}_a)(R_{b,c} - \bar{R}_b)}{\sqrt{\sum_{c \in I_{a,b}} (R_{a,c} - \bar{R}_a)^2} \sqrt{\sum_{c \in I_{a,b}} (R_{b,c} - \bar{R}_b)^2}} \quad (\text{公式 3})$$

式中： $\bar{R}_a$ 和 $\bar{R}_b$ 分别代表用户对项目的平均评分； $R_{a,c}$ 代表用户 a 对项目 c 的评分； $I_{a,b}$ 代表用户 a、b 对项目共同评分的集合； $I_a$ 和 $I_b$ 分别代表经用户 a 和 b 评分的项目集合。

### 2 基于项目的协同过滤推荐方法

Sarwr 教授<sup>[12]</sup>提出的基于项目的协同过滤技术的基本思想：计算已评价项目与待预测项目的相似度来寻找推荐对象的最近邻居。基于该算法的推荐精确度低，可对算法进行改进：首先基于项目特征属性矩阵计算项目间相似性，找出邻居项目的候选集，然后基于用户——项目评分矩阵计算目标项目与邻居项目候选集中项目的相似性，找出目标项目的最近邻居集。该算法利用项目的自身属性和用户评价的影响，虽然解决了数据的稀疏性和冷启动问题，但是对于新用户还是无法有效推荐。

### 3 基于模型的协同过滤推荐方法

基于协同过滤算法存在的数据稀疏和冷启动问题，提出了基于模型的协同过滤算法。它是根据数据库中的数据，建立相应的模型生成新的预测和用户偏好，模型有基于贝叶斯网络技术的模型、基于 SVD 奇异值分解技术的模型、基于关联规则的模型和基于聚类的模型<sup>[13]</sup>等。这种方法的关键在于如何将用户的最新信息实时的反馈给模型，从而提高推荐的精确度。

## 二 基于协同过滤的个性化课程推荐

### 1 推荐组成部分

选课推荐系统包含三大模块：学生建模模块、推荐课程建模模块、推荐算法模块。推荐系统的流程如图 2 所示。推荐系统把学生模型中的兴趣爱好、选课记录及学习状态和推荐课程模型中的特征属性匹配，同时使用合适的推荐算法进行计算筛选，为学生提供感兴趣的、新型的、热门的课程推荐。

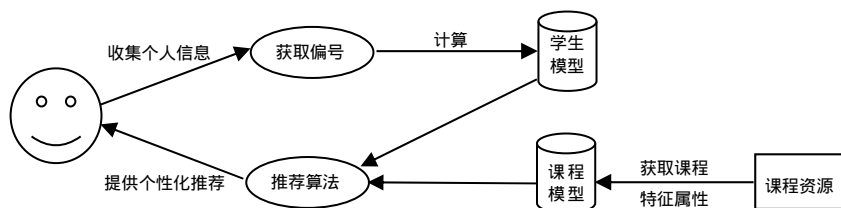


图 2 推荐系统模型图

### 2 基于协同过滤的课程推荐算法

推荐模块是推荐系统中的核心部分，对于推荐系统的性能和精确度起关键性的作用，提出如图 3 所示的基于协同过滤的个性化推荐算法流程：

- (1) 根据数据库中的学生信息表及选课信息，构建学生的特征向量，即学生模型；
- (2) 根据得到的学生模型和数据库中与课程相关的信息表，构建相应的课程特征向量，即课程模型；
- (3) 判断当前用户是新用户或老用户，若用户有选课信息记录则为老用户，否则为新用户；
- (4) 对老用户，若选课信息记录 $\geq 3$ ，采用基于用户的协同过滤算法为其产生推荐，否则采用基于项目的协同过滤推荐算法；
- (5) 对新用户，采用基于属性值偏好矩阵的最近邻推荐；
- (6) 最后在系统中增加排行榜课程推荐（用统计算法）和新型课程推荐（采用基于属性值偏好矩阵的新项目推荐）两个模块。

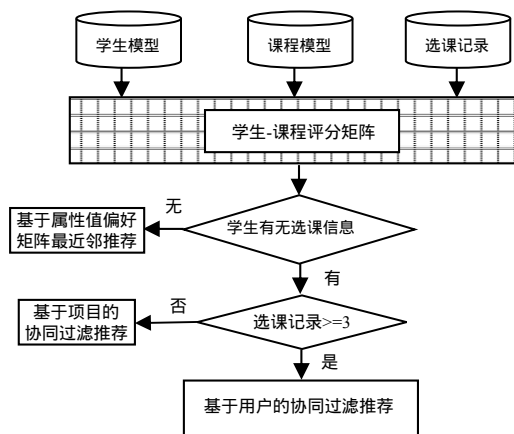


图 3 基于协同过滤的推荐算法流程图

推荐算法组合应用了多个协同过滤算法的优点,解决了算法的数据稀疏和冷启动问题。对于新用户采用基于属性值偏好矩阵的最近邻推荐,由于该算法采用的是离线,不能根据用户的兴趣动态更新,所以在该算法中加入了自适应算法定期更新用户的信息,提高算法的推荐准确性。

### 三 个性化推荐的选课系统

#### 1 系统框架

本系统采用 ASP.NET 应用程序典型的三层架构模式,即表现层通过统一的接口向数据访问层发送请求,业务逻辑层将请求按照一定的逻辑规则处理后进入数据库操作,然后将数据库返回的数据集合封装成对象的形式返回到表现层。

选课系统分为三个子系统:学生系统、教师系统和管理系统。根据选课的数据结构模型,在系统的数据库设计中,三个基本表是:学生表(Student)、教师表(Teacher)和课程表(Course)。

#### 2 功能模块

选课系统充分考虑到扩展学生的视野和提高学生的兴趣度等因素,划分了三大模块:个性化推荐、排行榜课程推荐、新型课程推荐模块。

##### (1) 个性化推荐模块

个性化推荐算法是根据前文所述的推荐算法流程执行的。其中基于属性值偏好矩阵的最近邻推荐的基本思路为:首先根据目标用户及其通过 SVD 降维获取的初始用户集中各用户的选课和评价记录,将其映射到各个项目的属性值上,生成用户的属性值偏好矩阵;其次对用户的属性值偏好矩阵进行相似度计算,得到目标用户的最近邻集合,并依据最近邻用户的选课记录实现课程推荐。

##### (2) 排行榜课程推荐模块

排行榜课程推荐中推荐课程分为两部分:一部分为将一定时间内课程被选的次数作为推荐依据;另一部分为众数法,通过计算每门课程的评价值,然后对评价值众数排序,选取前 M 门课程作为推荐依据。最后在选课推荐系统中的排行榜课程推荐为 M+N 门课程。

##### (3) 新型课程推荐模块

该模块是把新添加的课程推荐给合适的学习者,要做到既能扩展学生的视野,又能避免造成学生的反感。算法的基本思路是:(1)生成学生的属性值偏好矩阵;(2)根据新课程的加入及时获取该课程的属性值,保存在“新课程”表中;(3)将新课程的属性值与所有学习者的属性值偏好矩阵进行匹配,找出相似度最高的 Top-N 个学生;(4)当该课程被学生选修过,则从“新课程”表中删除该课程记录。为了不影响推荐系统的实时性,该算法采用离线的方式完成。

### 四 实验结果及分析

#### 1 实验数据集和度量标准

实验采用 ASP 开发了一套选课系统,后台数据库为 SQL Server 2008,评分制采用 5 分制。该系统收集了 100 种课程信息。推荐算法程序采用 Python 语言实现,其中 SVD 分解采用 MATLAB 计算。实验对象为 50 位来自教育技术学、计算机科学与技术两个专业的学生。实验要求每个学习者至少对 7 种课程给出评价。系统一共收集了 3428 条评分数据和 120 条选课记录,以此作为实验集。随机从每位学习者的评分数据中选取 5 条记录作为测试集,剩余为训练集。

推荐系统的推荐质量指标主要包括算法的复杂度、算法的 MAE 值和准确 (precision) 率/召回率 (recall), 还有其他指标, 如用户满意度、覆盖率、多样性、新颖性、惊喜度、信任度、健壮性。

其中准确率/召回率两个参数是评测给用户的推荐列表的精确度。召回率是指用户——项目评分记录在最后的推荐表中所占的比例, 而准确率是指在最终的推荐表中, 有多少是发生过的用户——姓名评分记录, 假设  $R(u)$  是用户在测试集上的行为列表,  $T(u)$  是根据用户在训练集上的行为给用户做出的推荐列表, 其准确率/召回率的计算公式如公式 4 和公式 5 所示<sup>[9]</sup>:

$$Recall = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |T(u)|} \quad (\text{公式 4})$$

$$Precision = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u |R(u)|} \quad (\text{公式 5})$$

MAE (Mean Absolute Error) 是通过绝对值计算预测评分与实际评分之间的偏差, 偏差值越小, 预测越准确。假设用户对项目的评分集合为  $\{Q_1, Q_2, \dots, Q_n\}$ , 而用户对未评分项目的预测评分集合为  $\{\hat{Q}_1, \hat{Q}_2, \dots, \hat{Q}_n\}$ , 则 MAE 用公式 6 表示为:

$$MAE = \frac{\sum_{i=1}^n |Q_i - P_i|}{n} \quad (\text{公式 6})$$

MAE 值越小, 推荐质量就越高。

## 2 实验结果及分析

实验将本文的改进算法与其他推荐算法 (热门项目推荐、随机项目推荐、基于项目的协同过滤推荐) 进行分析比较, 结果如表 1、表 2 所示。

表 1 两种算法的准确率/召回率和覆盖率

算法	准确率	召回率	覆盖率
User-attriHostItem	2.94%	6.00%	3.85%
HostItem	2.36%	4.81%	0.018%

表 2 不同算法的对比分析

算法	覆盖率	准确率	召回率	流行度
ItemCF	33.71%	2.56%	9.44%	2.9119
HostItem	0.0299%	1.18%	4.36%	7.1277
Random	84.18%	0.000985%	0.00305%	0.9878
Item-attriCF	34.44%	6.98%	25.75%	1.7086

实验表明, 基于用户属性和基于项目属性的协同过滤改进算法的准确率/召回率要比其它算法相对较高。

本文还采用该算法与基于用户的协同过滤算法进行比较, 并将推荐集的项目数设为 10, 最近邻居数分别取 2、4、6、9、11, 然后分别采用这两种算法对 MAE 值进行比较, 结果如图 4 所示。

结果表明基于用户、项目的属性和属性值偏好矩阵的协同过滤算法的 MAE 值比较小, 这是

由于预测值是根据最近邻居预测的,降低了预测值和实际评价之间的误差,而 MAE 值越小评分预测越准确,推荐质量越高。

本文的改进算法在最近邻居数取 9、11 时求得的推荐集中,发现了没有被学习者选修过的和没有被评过分的课程,这表明该算法解决了项目的冷启动问题,即实现了新课程推荐。

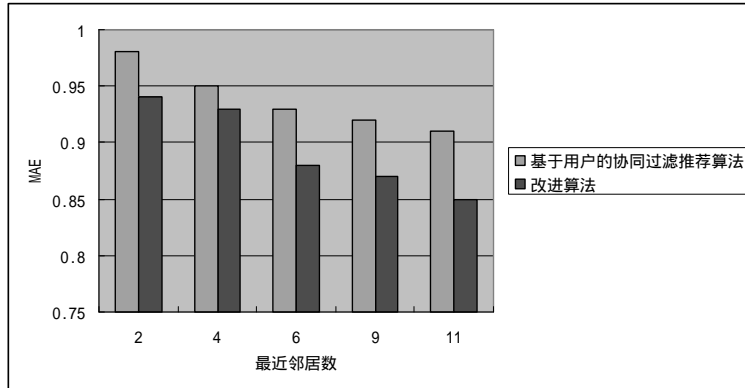


图 4 推荐算法的 MAE 比较

## 五 小结

针对选课系统无法满足学生个性化的选课推荐的缺陷,本文提出了一种基于用户、项目及属性值矩阵的协同过滤推荐方法,并采用该方法设计了一个基于协同过滤技术的个性化推荐的选课系统,能有效地进行个性化课程的推荐,减少学生选课的盲目性,从而提高选课效率和学生选课的满意度。本文主要阐述了如何解决协同过滤算法的数据稀疏性和冷启动问题,而且实现了选课系统的个性化推荐,尚未对选课系统中的数据安全性和用户长时短时兴趣的区分作进一步的研究,这些内容将在下一步的研究中进行探讨。

## 参考文献

- [1]王国霞,刘贺平.个性化推荐系统综述[J].计算机工程与应用,2012,(7):66-76.
- [2]王永固,邱飞岳,赵建龙,刘辉.基于协同过滤技术的学习资源个性化推荐研究[J].远程教育杂志,2011,(3):66-71.
- [3]Resnick and Varian. Recommender systems[J].Communications of the ACM,1997,(3):56-58.
- [4]于建东.学分制条件下我国高校选课制中的问题与对策[J].文史博览(理论),2009,(6):68-70.
- [5]曾春,邢春晓,周立柱.基于内容过滤的个性化搜索算法[J].软件学报,2003,(5):1000-1003.
- [6]Goldberg K,Roeder T,Gupta D,Perkins C.Eigentaste:A constant time collaborative filtering algorithm.Information Retrieval,Communications of the ACM,1997, (3):77-87.
- [7]Shardanand U,Maes P.Social information filtering:Algorithms for automating “Word of Mouth”,In:Proc. of the Conf.on Human Factors in Computing Systems.New York:ACM Press,1995:210-217.
- [8]Konstan JA,Miller BN , Maltz D,Herlocker JL,Gordon LR,Riedl J.GroupLens:Applying collaborative filtering to usenet news.Communications of the ACM,1997, (3):77-87.
- [9]项亮.推荐系统实践[M].北京:人民邮电出版社,2012:44-59.

- [10]宋雅婷,徐天伟.基于用户兴趣的个性化推荐技术的研究综述[J].云南大学学报,2012,(5):1000-1003.
- [11]Xu Bin, Liu Sai, Lishan, et al. Information retrieval technique based on genetic algorithm [J]. Computer Engineering, 2004,(9): 74-75.
- [12]Yu Li,Liu Lu,LiXuefeng.A hybid collaborative filtering method for multiple-interests and multiple-content recommendation in E-commerce[J].Expert Systems with Applications,2005,(1):67-77.
- [13]解亚萍,牛广文.用户聚类的电子商务推荐系统研究[J].兰州工业高等专科学校校报,2009,(3):11-13.
- [14]李聪,梁昌勇.基于属性值偏好矩阵的协同过滤推荐算法[J].情报学报,2008,(6):884-890.

### Research on the Elective System with Personalized Recommendation Based on Collaborative Filtering

XU Tian-wei<sup>1,2</sup>    SONG Ya-ting<sup>2</sup>    DUAN Chong-jing<sup>2</sup>

(1. Institute of Education Sciences, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China;

2. School of Information Science and Technology, Yunnan Normal University, Kunming, Yunnan 650500, China;)

**Abstract:** Problems of lacking in individualized curriculum recommendations and inefficiency exist in current course selection systems of institutions of higher education. In allusion to these limitations, this paper presents a novel collaborative filtering algorithm based on the project, user and attribute-value matrix through analysis and study of personalized recommendation technology. The proposed algorithm has been successfully applied to the elective system. Experimental results indicate that the proposed approach can solve cold-start technology in personalized recommendation algorithm, improve the related indicators significantly, and achieve a personalized recommendation and new courses recommendation.

**Keywords:** course selecting; personalized recommendation; collaborative filtering

---

\*基金项目：本文系国家自然科学基金项目“基于 e-Science 的民族信息资源融合与语义检索研究”（项目编号：61262071）的研究成果。

作者简介：徐天伟，云南师范大学信息学院，博士研究生，研究方向为教育信息化、软件工程方面研究。

收稿日期：2013 年 12 月 13 日

编辑：小西