**基于协同过滤的个性化选课模式研究与实践[[1]](#footnote-1)**

孙彦超

(北京信息科技大学 教务处，北京 100192)

摘 要 针对协同过滤算法存在数据稀疏性问题及新用户问题,根据选课系统的具体情况及特殊性，比如，优秀学生可以按相似度高的邻居评价推荐，对于成绩较差的学生应参照优秀学生的选课情况对其推荐。本文对协同过滤算法进行改进，在学生对课程评价矩阵的基础上，抽取出一部分具有指导意义的信息，称作“优秀学生”评分矩阵，该矩阵由选课系统中所有评分数据过滤生产，代表了系统所有学生的评价信息，并随着系统中评价数据的变化而改变，利用“优秀学生”评分矩阵，根据学生的学院、专业、性别等属性，计算目标学生与“优秀学生”间的相似度；进而生成最近邻居集合；最后根据最近邻居对其生产高质量的推荐。改进后的算法在北京信息科技大学选课系统中进行实验,实验结果表明,改进后的算法在推荐效率及准确度上有明显的提高。

关键词 协同过滤；个性化选课模式；稀疏性；评分矩阵；最近邻居

**中图分类号:** TP3

**Research and practice-based collaborative filtering personalized elective mode**

Sun Yanchao

(Educational administration office, (Beijing information science & technology university, Beingjing 100192,China)

1. Collaborative filtering algorithms exist for data sparsity problem and issue a new user, depending on the circumstances and the special nature of elective system, for example, the best students can recommend a similar high evaluation neighbor, for the poor performance of students should refer to the outstanding student enrollment circumstances of its recommendation. In this paper, collaborative filtering algorithm is improved on the basis of student course evaluation matrix, part of the information extracted instructive, called "outstanding student" scoring matrix that filters all ratings data produced by elective system, on behalf of the evaluation information system for all students, and with the changes in the data evaluation system is changed, the use of "outstanding student" scoring matrix, according to the student's college, professional, gender and other attributes, compute the similarity target students with "outstanding student" among ; thereby generating nearest neighbor set; final recommendation based on the nearest neighbors of their production of high quality. The improved algorithm experiments in Beijing Information Science and Technology University elective system by experimental results show that the improved algorithm on the recommended efficiency and accuracy has been significantly improved.
2. Collaborative filtering; personalized enrollment patterns; sparsity; scoring matrix; nearest neighbor

# 0引 言

随着高等教学改革的不断深化以及信息化工作的快速推进，许多高校采用学籍管理系统进行教学管理，利用课程评价提高教学质量。依据课程的教学大纲和教学目标要求，课程评价采集学生对教师在教学过程中的各种教学活动效果的评价。同时收集教师同行或者专家对课程教学活动的评价，以此作为该门课程任课教师的教学质量的重要参考依据，最终反馈给学校和任课教师，帮助任课教师扬长避短，相应调整教学活动并提高教学质量。一般来说，大部分高校的教学质量评价信息主要用来指导教师教学活动，没有挖掘出评价信息的深层价值。例如，学生的评价信息反映了学生对课程的兴趣度，反映了学生对选课的个性化需求。另外,当前的高校选课模式受传统的学年制观念影响,没有充分发挥如何培养高素质应用型创新人才的功能,使得选课制度无法满足了学生的个性化学习需求。同时学生选课指导工作没有落实到位，学生选课时带有很大的盲目性，特别是新生，面对大量的开课信息表现的不知所措。因此，对当前高校选课制度的改革，如何正确引导学生选课，如何调动学生的积极性，对高校教学改革具有重要意义。

目前，高校学籍管理系统中存储了海量的教学评价信息，为了充分利用这些评价信息，大部分高校使用数据挖掘技术，利用关联规则对其进行统计分析，分别从学生、教师、课程等维度进行分析[1]。一部分高校在学生选课系统中应用基于关联规则技术为学生推荐选课信息，由于支持度和置信度设置不合理，再加上算法效率低、耗时长，严重影响了推荐选课的效果。同时，关联规则技术把所有学生同样对待，没有考虑到不同专业及不同性别等特性学生的个性化需求。目前在推荐系统中，协同过滤算法凭借算法简单实现容易、对数据对象依赖性较低及推荐准确率高等特性被广泛采用。笔者将该算法引入学生选课系统中,针对选课系统的具体特性,在生成用户评价矩阵时对其算法进行改进,从而更好地为学生选课产生具有指导意义的推荐活动。

# 1协同过滤算法介绍

协同过滤算法(Collaborative Filtering recommendation)的概念由Goldberg等研究人员于1992年提出[2]。该算法的先决条件是：1)用户之间存在相似的兴趣，用户的兴趣度可以根据对物品的评分来衡量；2)用户的兴趣相对稳定，可以依据用户历史兴趣产生推荐。协同过滤算法归纳为三步：1)生成用户评分矩阵；2)分析用户相似度计算最近邻居集；3)生成推荐结果[3]。

协同过滤算法凭借其算法简单且容易实现、对数据对象依赖性较低、推荐准确率高等特点得到主流推荐系统广泛采用[4]。该算法根据相似度将用户分为不同的群体，根据同一类的用户兴趣对目标用户进行推荐。但在学生选课系统中，如果将学生分成不同的群体，势必会产生成绩好的学生为一类，成绩差的学生为另一类，根据同类兴趣度产生推荐选课，将会在学生选课方面造成两级分化，进而影响个性化推荐系统设计的初衷。针对这一问题，本文对协同算法进行改进，在生成学生评分矩阵后，对用户评分进行筛选，生成有参考意义的“优秀学生”评分矩阵，在“优秀学生”评分矩阵上根据学生所学专业及个人属性分析相似度，依据相似度计算最近邻居，从而为学生推荐选课信息。

# 2协同过滤算法改进

协同过滤算法能够为用户提供个性化推荐，但该算法在个性化选课系统中存在一系列问题[5]：1)依据所有学生对课程评价矩阵，对于学习成绩较差的学生来说，根据评分相似度推荐选课，推荐的质量较低。比如，某一学生属于“差等生”，但该生希望跟“优等生”一样被推荐选课，该算法往往给推荐一些“差等生”所感兴趣的课程；2)学生和课程数量大，而且不断增长，每个学生在校期间最多选择上百门课程，学生对课程的评分数据占矩阵中数据量的很少一部分，因而造成学生对课程的评分十分稀少，依据生成的稀疏矩阵产生的推荐质量较低；3)每学期会有几千名新生数据和若干课程导入选课系统，这些学生和课程没有评价信息。对学生新用户来说，根据其评价信息计算其最近邻居集，将无法为其产生推荐课程；对新课程来说，没有被评价信息，无法把自己推荐给学生，这就是通常所说的冷启动问题[6]。

针对以上三方面问题，本文对原有算法进行改进，改进思路是在协同过滤算法生成的学生对课程评分矩阵上，对该评分矩阵进行抽取，抽取原则是利用学生成绩库，统计出课程学习成绩为良好学生作为有效用户，保留有效用户的对应课程评分数据。在剩余的评分矩阵上通过挖掘算法，抽取并生成“优秀学生”评分矩阵，参照学生属性（专业、性别、偏好）在评分矩阵上生成最近邻居集，从而为目标学生进行推荐选课。所谓的“优秀学生”并不是具体的某一学生，是依据成绩库和学生评分数据抽取出来的一部分具有代表性的学生，兼顾了一般学生和优秀学生的兴趣。首先，在生成“优秀学生”矩阵的抽取过程中，每个学生的评分都参与抽取过程。因此，“优秀学生”矩阵能够准确地反应所有学生的兴趣。其次，过滤掉一部分不具备指导意义的评分数据，同时对新进入系统的学生用户，可以根据本专业和自己偏好相似的优秀学生评分，作为新用户学生的评分，为其产生推荐，即能为所有学生产生具有指导性的选课推荐，又能解决传统协同过滤算法经常遇到的稀疏性和冷启动问题。改进后的算法推荐步骤如图1所示。

生成学生评分矩阵

抽取生成“优秀学生”评分矩阵

参照学生专业性别爱好计算最近邻居集

生成推荐结果

图1改进后协同过滤算法步骤

改进后的算法可以分为抽取“优秀学生”矩阵和产生推荐课程两个步骤。

## 2.1抽取并生成“优秀学生”矩阵

该步骤主要目的是利用学生评分矩阵及成绩库生成“优秀学生”矩阵。首先根据成绩库统计每门课程成绩取得良好或优秀的学生，抽取出这部分学生对该门课程的评分数据。依据统计出的数据从学生评分矩阵中过滤掉一部分不具备指导意义的评分记录。利用保留的学生评分记录成生成学生评分矩阵R，利用该矩阵首先建立一个学生s对课程c的评分预测公式：

** (1)

式(1)中，为学生s对课程c的测评分；和分别为目标学生和“优秀学生”对课程的平均评分；对应矩阵的第s行c列，即学生s对课程c的评分值(R矩阵中的元素)；N为优秀学生集合；为目标学生和优秀学生的相似度。改进后的算法计算采用式(2)计算。

** (2)

通过式(2),使用学生对课程评价矩阵R作为训练集，可以利用式(3)计算预测值和实评分值的误差。

**(3)

利用预测公式(1)作为生成公式，把学生评价矩阵R作为训练集进行多次迭代，利用公式(3)控制精度，最终生成一个具有指导价值的“优秀学生”评分矩阵。

## 2.2产生推荐课程

该步骤主要目的是预测目标学生s对待选课程c的评分，选择预测值最高的多门课程最为推荐结果显示给选课学生。其中目标学生与“优秀学生”之间的相似度计算算法非常重要，算法的好坏直接影响到推算的准确度，由于不同专业学生选课侧重点具有较大差异，另外学生的性别爱好对学生相似度具有较大影响。在计算用户相似度时引入学生特性相似度式（4）对公式(2)进行改进，

** (4)

式（4）中为学生特征相似度；为学生专业相似度；为学生性别相似度；为学生兴趣相似度；，分别为相关系数。进而生成式（5）可以更加准确的计算学生相似度。

**(5)

利用式(5)生成目标学生的最近邻居集，根据目标学生的最近邻居集，利用式（6）进行推荐。

**(6)

式（6）中，M为目标学生s的最近邻居集合；表示邻居j对待评课程c的预测评分；表示邻居j对所有课程评价的平均值。通过对推荐公式（6）预测其对待选课程的评分，选择评分较高的N门课程最为推荐课程，对选课学生提供推荐服务。

# 3算法应用与分析

实验数据选用北京信息科技大学2007级、2008级、2009级及2010级四届本科生大学期间所选课程及评价数据，共11252名学生、1870门课程。由于学校对学生采取了有效的课程评价管理措施，每门课程评价学生占选课学生的80%以上,因此评价数据能够很好反应本校学生对课程的评价信息。同时将学生评分数据集分为训练集和测试集两部分，其中训练集和测试集分别占总数据的85%和15%。进而采用平均绝对偏差MAE来评价改进后算法的推荐质量，MAE越大表明推荐质量越差。根据式(7)可计算平均绝对偏差为:

** (7)

利用式（7）测试改进算法前后平均绝对误差时，将目标邻居集合从5个增加到25个，实验结果如图2所示。

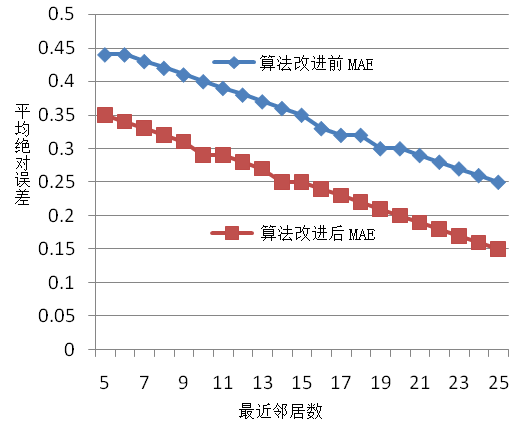


图2改进算法前后平均绝对误差

从图2可得，改进后的算法产生的推荐平均绝对误差均比改进前小，也就是说，改进后的算法从平均绝对误差角度分析能够提供高质量的推荐效果。

为了验证改进前后算法在查准率式（8）和误判率式（9）方面的效果，对100名学生问卷调查，其中100名学生来自25个专业，每个专业抽取4名有代表性的学生，综合学习成绩分别为优秀、良好、中等和较差，由于基本上每个专业及各类层次的学生都考虑到，因此选取对象具有代表性,针对改进前后算法对其产生推荐结果的查准率及误判率进行问卷调查。统计改进前后查准率和误判率。

** (8)

** (9)

式(8)及式(9)中,H为推荐的正确数目;N表示推荐的总个数。查准率及误判率统计结果如图3所示。

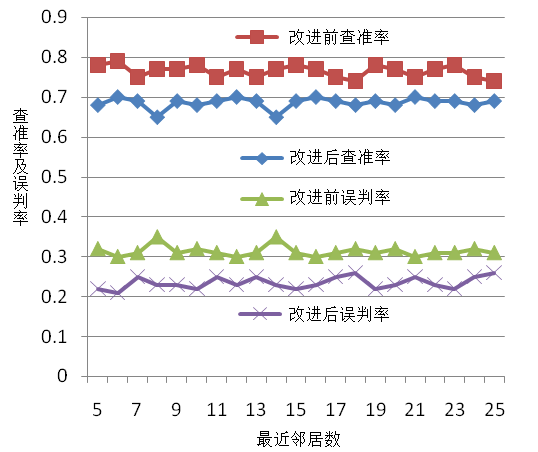


图3 算法改进前后查准率和误判率实验结果

由图3可得，改进后算法查准率均比原算法准确率高，能够为学生选课提供较好的推荐服务。

# 4结束语

针对协同过滤算法在个性化选课系统中对“差等生”推荐质量低，同时学生评分数据在评分矩阵中存在稀疏性和冷启动问题，本文对协同过滤算法改进，首先对学生对课程评价矩阵进行信息过滤，通过抽取“优秀学生”数据，进而生成“优秀学生”评价矩阵，在计算目标学生相似度时，引入学生特征相似度函数计算相似度，生成最近邻居集，并依据最近邻居集为目标学生提供推荐服务。改进后的算法在北京信息科技大学选课系统中测试，实验结果表明，改进后的算法平均绝对误差小于原算法，并且准确率高于原算法。从而说明改进后算法能够提供更好的推荐质量。

参考文献

[1]铁治欣,陈奇,俞瑞钊,等. 关联规则采掘综述[J].计算机应用研究, 2000,4(1):122-124.

[2]邢政权. 数据仓库与数据挖掘技术在教学管理系统中的研究与应用[D] .南京：南京理工大学出版社， 2004.

[3]田保慧. 正相关频繁项集的挖掘算法[J].华北水利水电学院学报,2008,6(4):101-102.

[4]李金娟,王卫锋.基于FP-growth算法在学生成绩中的关联规则分析[J].巢湖学院学报,2007,12(6):212-214.

[5]KonstanA, MillerB, et al. GroupLens: Applying Collaborative Filtering to USENET News[J]. Communication of the ACM,2011(3).

[6]Shi Hua.Item-Based and User-Based double clustering Collaborative Filtering Recommendation Algorithm[D].Changchun:Dongbei Normal Univesity,2009:2-3.

[7]李雪峰，刘鲁，张墨．基于协同过滤的在线拍卖商品推荐[J]．计算机工程,2006,8(3):18-20．

[8]郫艳红，邓贵仕．协同过滤的一种个性化推荐算法研究[J]．计算机应用研究，2008,5(3):139-141．

[9]赵亮，胡乃静，张守志．个性化推荐算法设计[J]．计算机研究与发展，2002,31(2):615-618．

[10]SUN Xian-hua．Research of sparsity and cold start problem in collaborative filtering[D]．Hangzhou：Zhejiang University，2005,8（3）:73-75.

[11]杨春蓉．高校公选课网上选课系统的设计与实现[D]．上海:华东师范大学出版社，2007．

[12]陈月英，宗平，庄卫华．高校选课系统中的公平算法及其研究[J]．计算机工程与应用，1998,11(3):316-319．

[13]谢丹霞，黄争舸．信息时代高校选课系统的智能化构建[J]．高等农业教育,2006,3(2)：63-65.

[14]董坤.基于协同过滤算法的高校图书馆图书推荐系统研究[J]．知识组织与知识管理，2011,21(1):910-912.

[15] 杨永权.基于协同过滤技术的图书推荐系统研究[D].广州:华南理工大学出版社,2013.

[16] 林鸿飞,杨志豪,赵品.基于内容和合作模式的信息推荐机制[J].中文信息学报,2005,25(3):715-716.

[17] Tao Jun,Zhang Ning.Collaborative Filtering Algorithm Based on Interest-Class[J].Computer Systems&Applications,2011,19（2）:183-185.

[18]Wang Hui-min,Nie Gui-hua.Research on Collaborative Filtering Algorithm Based on Fusing User and Item’s Correlative Information[J].Journal of Wuhan University of Technology,2007,11（2）:110-115.

1. 收稿日期：2014-12-16

   基金项目：北京信息科技大学2014年教改项目（2014JG31）

   作者简介：孙彦超（1978-），男，河南南阳人，博士研究生，工程师 [↑](#footnote-ref-1)