ISSN 1006 - 7167
CN 31 - 1707/T
实验室研究与探索
RESEARCH AND EXPLORATION IN LABORATORY
第 36 卷 第 5 期 2017 年 5 月
Vol. 36 No. 5 May 2017

校园二手书交易网站推荐引擎的设计与实现

丁 一， 沈至榕， 谢颖华
(东华大学 信息科学与技术学院，上海 201620)

摘 要: 大学二手书交易同一般网络购物不同，直接应用普通推荐算法设计推荐引擎不能很好地进行推荐。为了使大学生二手书购物网站能够更好地为用户推荐书籍，提高推荐结果的准确性及推荐效率，在原有推荐模块的基础上针对大学生用户的特殊性提出了一种适用于二手书购物网站的推荐算法模块优化推荐引擎。该算法模块有以下特性: ①在基于物品推荐算法的基础上，结合书籍评分、新旧等因子优化推荐结果; ②为方便用户购书新增打包推荐功能，使用户可以一键购买多本所需书籍; ③根据用户群体购书可预测的特性加入了基于时间节点的推荐算法，即按时得需，按需推荐。

关键词: 推荐算法; 评分机制; 新旧因子; 打包推荐; 时间节点

中图分类号: TP 39 **文献标志码:** A **文章编号:** 1006 - 7167(2017)05 - 0148 - 06

Design and Implementation of Recommendation Engine Based on University Campus' Used Book Shopping Network

DING Yi, *SHENG Zhirong*, *XIE Yinghua*
(School of Information Science and Technology, Donghua University, Shanghai 201620, China)

Abstract: College used book trade is different from the general online shopping. Recommendation engine applied common recommendation algorithm cannot be very good to recommend. In order to make students' used book shopping site better to recommend books to users, to improve the accuracy of the results and the efficiency of recommendation, based on the original recommendation module, the paper proposed a suitable mode for shopping website of used books recommendation engine with an optimal recommendation algorithm module. The new algorithm module has the following characteristics: (1) Based on the items of the recommendation algorithm, books' scores and "old and new" factors are taken into consideration to optimize recommendation results; (2) For the convenience of users purchase, a new packaged recommendation function is added, so users can buy the required books by pushing one key; (3) According to the fact that the needs of students can be predicted, the recommendation algorithm was designed based on time nodes.

Key words: recommendation algorithm; scoring mechanism; "old and new" factors; packaged recommendation function; time-based

0 引 言

随着可持续发展理念的进一步加强，为了节约书本资源，二手书交易在高校蓬勃发展。在信息化时代的背景下，二手书互联网交易平台应运而生。二手书购物网站出售的书籍种类繁多，涉及到教材、课外读物等多个领域，此外，教材类书籍又包含学校多个专业的不同教材，简单的站内搜索引擎[1]已不能很好地满足用户需要，因此二手书购物网站同样面临着信息过载[2]的问题，个性化推荐引擎的应用与创新便更加具有价值。而大学二手书交易同一般网络购物不同，大

收稿日期: 2016-08-22

作者简介: 丁 一(1994 -)，男，河北张家口人，本科生，主要研究方向: 数据挖掘。Tel.: 18818010121; E-mail: 200188881@qq.com

通信作者: 谢颖华(1972 -)，女，上海人，硕士，副教授，主要研究方向: 大数据。Tel.: 13701773541; E-mail: yh_xie@dhu.edu.cn

学二手书交易具有以下特点: ①用户群体的特殊性,均为在校大学生; ②购书需求的特殊性及可预测性,一般为科本教材、辅导读物等; ③大学二手书与普通新书交易不同,具有新旧、时间等影响因子[3-4]。因此,直接应用普通推荐算法设计推荐引擎不能很好进行推荐。推荐算法是推荐引擎的核心内容,也是决定推荐效率以及推荐品质的关键。推荐的个性化以及合理化是衡量推荐算法的重要指标。本文结合实际情况合理设计与改进推荐算法,完善推荐引擎,使得网站的用户能够得到及时、有效且相对满意的推荐结果。

1　一般推荐算法的应用及问题

1.1　基于物品的推荐算法的应用

第一部分: 基于物品的推荐算法(Item-Based Method)[5],该算法是为物品分类标签,将与用户曾经购买过的物品属性相同或相似的另一种物品推荐给该用户。针对本次应用对象——大学生,本文将书籍分类为小说、课本、辅导书等,并且在各大类的每本书上加上相关标签。例如《模拟电子技术基础同步辅导及习题全解》这本书,其标签有①作者: 于登峰,边文思; ②类别: 辅导书; ③适用课程: 模拟电子技术基础④专业大类: 电气信息类等。如果该用户之前买过《模拟电子技术基础》这本书,便将该书推荐给该用户。同时,基于用户具有特殊性[6-7]的事实,根据用户所在学院,专业以及年级的不同,对其进行相应的推荐,使推荐结果更具针对性,实效性。

第二部分: 基于物品的协同过滤算法(Item-Based Collaborative Filtering Recommendation Algorithms)[8]。本文针对书籍的标签机制[9-10],将该算法应用于此次推荐算法中。在两本书之间有两个以上的相同标签的情况下,根据以下公式计算两个物品之间的相似度[11-12]:

$$P_{uj} = \sum_{i \in N(U) \cap S(j,K)} w_{ji} r_{ui} \qquad (1)$$

式中: $N(u)$ 是含有用户喜欢的物品的集合; $S(j,K)$ 是和物品 j 最相似的 K 个物品的集合,这里根据书籍标签相似度取 K 为 10,即取与书籍 j 相似度排名前十的书籍的集合; w_{ji} 是书籍 j 与书籍 i 的相似度; r_{ui} 是用户 u 对物品 i 的兴趣。利用该公式以及标签机制,可以在用户数据缺乏的时候对用户做出尽可能准确对推荐。

1.2　一般推荐算法未能解决的问题

以上算法是基本算法,利用物品之间对的相似度以及相互之间的联系(课本与该课本辅导书的关系)对用户进行推荐,这样的好处是简单高效。由于大学校园二手书交易的特殊性,这样的算法有以下待解决的问题: ①推荐结果未考虑二手书交易特有的影响因子——新旧程度; ②重复推荐,即推荐给用户其已有或不需要的物品,例如学生只需要 1 本关于《模拟电子技术基础》的辅导书,而推荐给他内容相近的 3 本辅导书,那么对于用户而言推荐后仍然需要筛选,降低推荐效率; ③用户购书需一本一本查找选择,不能使用户方便高效地购物; ④推荐结果具有滞后性,不能预测用户购物需求,即完全依赖用户购买记录,没有用到用户基本属性的预测作用。⑤结合标签机制的协同过滤算法面临的问题是标签数量有限导致相似度处于同一等级的书籍数量较多,会使推荐结果繁多。

2　推荐引擎设计

2.1　推荐引擎的结构

2.1.1　推荐系统的结构功能[13]

(1) 数据收集和存储模块。UI 系统负责给用户展示网页并和用户交互。网站会通过日志系统将用户在 UI 上的各种各样的行为记录到用户行为日志中(见图 1)。

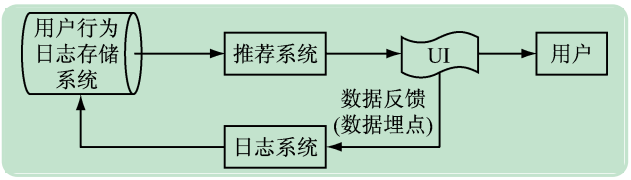

图 1　推荐引擎结构图

(2) 推荐引擎模块。它是推荐系统的最核心部分,采用的推荐技术决定着推荐系统的性能优劣(见图 2)。本文所采用的推荐函数一共有4个,分别对应

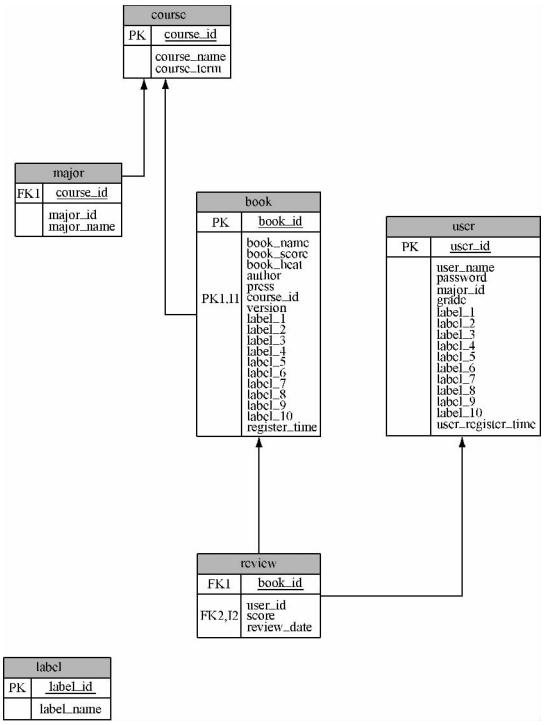

图 2　推荐引擎模块结构

下文结合新旧因子的推荐、基于评分机制的推荐、打包推荐、基于时间节点的推荐算法的应用。

2.1.2 数据库设计

数据库用于保存和管理用户行为数据和所售商品的特征数据，是推荐引擎的数据存储模块。本文针对此次应用实例将数据库设计如图 3 所示。

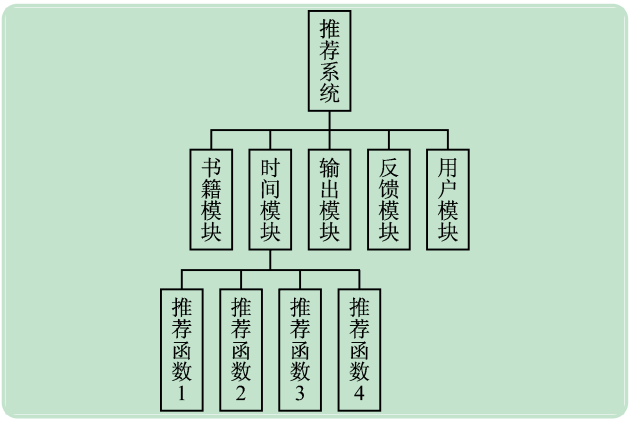

图 3 数据库 ER 图

2.2 推荐算法模块的设计

推荐算法模块是推荐引擎的核心部分，该部分所采用的推荐技术直接决定推荐引擎推荐效果的品质。本文针对大学校园二手书交易平台的实际情况设计完善推荐引擎的推荐算法模块，以解决上文分析的待解决的问题。

2.2.1 结合新旧因子的推荐

二手书交易不同于新书购物，具有特定影响因子：新旧程度。可以为书本增添新的标签用来记录书本新旧程度。书籍上架前，由工作人员根据一定的标准将书籍划分为 1 ~4 个等级，其中等级 1 为九成新；等级 2 为八成新；等级 3 为七成新；等级 4 为六成新及以下。等级对应书籍品质参阅京东网对二手书分级标准，结果见表 1。

表 1 二手书新旧程度对应等级表

新旧等级	书籍品质
等级 1	图书保存完好。图书品相项均没有问题
等级 2	保存较好。图书品相常见问题[2] 项均没有问题。常见问题[2] 可有轻度存在，但不超过图书总面积的 5%，不影响文字阅读和图书整体美观
等级 3	保存一般。图书品相常见问题[2] 项均没有问题。常见问题[2] 可有轻度存在，但不超过图书总面积的 10%，基本不影响文字阅读和图书基本美观
等级 4	保存不佳。书的主体部分不缺，图书品相常见问题[2] 可有轻度存在。常见问题[2] 一定程度存在，但不超过图书总面积的 30%，一定程度影响文字阅读和图书美观。图书破损较严重或者图书重要项目有缺失，需要用文字在商品描述中说明

注：常见问题[1] 残缺、粘连、水渍、变形、虫蛀、修补；
常见问题[2] 墨迹、污损、签章、磨损

将二手书新旧等级赋值变量 I_{cr} 用于推荐价值计算。

2.2.2 基于评分机制的推荐算法

针对原有基于物品推荐算法的重复推荐的问题，本文采用建立评分机制[14]，并结合书籍热度对推荐结果进行优化的方法（见图 4）。在本文中热度指的是一本书被评分的次数。为了避免因出现一本书虽然评分人数较少但却评分很高这种情况，导致其推荐次序靠后的问题，本文采用对评分结果和热度进行加权相加来计算推荐价值的大小。

针对基于物品的协同过滤算法计算相似度式（1）做出以下调整：$N(u)$ 表示所有含有用户喜欢的标签的书的集合，r_{ui} 的值：当用户 u 对书籍 i 评分过为“1”，否则为“0”；并增添 w_{ji} 的求解公式[11-12]：

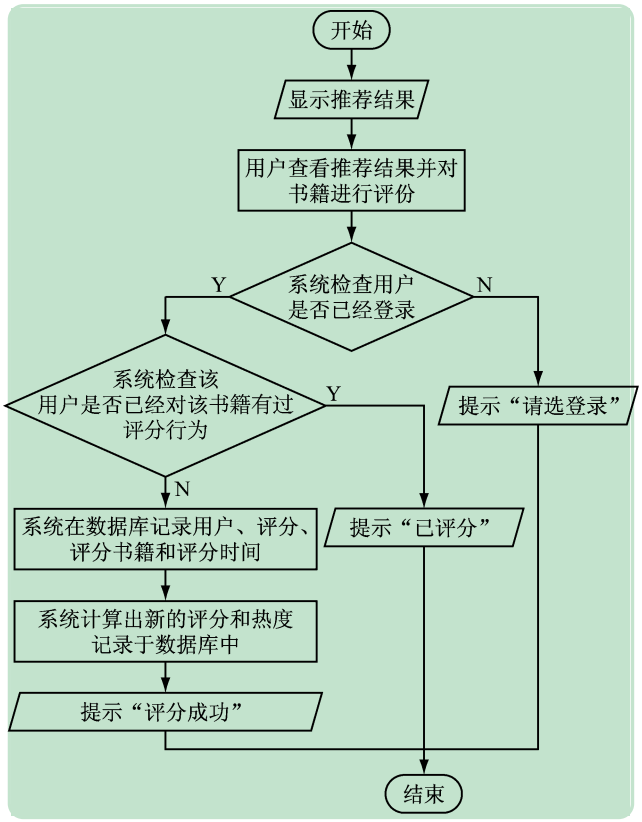

图 4 基于评分机制的推荐算法流程图

$$w_{ji} = \frac{M_{ij}}{\sqrt{I_{ui} \cdot J_{uj}}} \tag{3}$$

式中：M_{ji} 为 M 对书籍 j 和书籍 i 评分过的用户；I_{ui} 对书籍 i 评分过的用户数；J_{uj} 为对书籍 j 评分过的用户数。

推荐价值计算公式如下：

$$P_i = aS_i + bH_i + 1/I_{on} \tag{4}$$

式中：P_i 为书籍 i 的推荐价值；a、b 为比例系数，本文采用 $a = b = 0.5$；S_i 为书籍 i 的评分值；H_i 为书籍 i 的

热度(即被评分次数),I_{on}为书籍新旧等级。

加入评分机制后,可以更加客观地对书籍的推荐价值进行排名,基于“择优推荐”的基本构想,便可以大大提高推荐质量;同时为对推荐数量进行限制,进一步优化推荐结果;结合热度的加权算法可以提高推荐结果的准确度。

2.2.3　打包推荐功能

在原有推荐算法模块的基础上,结合用户群体的特质进行改进,加入打包推荐的功能。通过调查,用户购书高峰期一般为学期末和开学前,用户主要购书需求是对下一学年所开课程的课本。二手书购物网的书籍一般是按本出售的,所以用户面临的问题是①一学期所开设的课程内容需要到学校教务网查询;②挑选课本时需要按课程表一一查找,费时费力。针对以上问题在原有的算法基础上加入了“打包推荐”的理念,在后台服务系统中将学院不同专业各个学期所开设的课程以及课程所需课本(不含辅导书)记录下来,将其中所设计的图书打包为“XXX 专业大 X 第 X 学期课本包”,将这一选项添加到推荐结果中去。

增加该功能的优点是可以将多本图书一步推荐,同时解决了学生不清楚课本的种类的问题,免去了学生查询教务网、按课表一一查找的繁琐,提高了推荐效率和用户使用体验。该项功能的缺点是需要保证后台数据的正确性与实时性,即专业课程安排要正确,并且当学校课程安排有变化时要及时更新打包推荐的内容。同时要确保推荐内容与推荐对象的匹配正确,即推荐内容要与用户的专业和学期相对应。

2.2.4　基于时间节点的推荐算法

通过分析得出购书高峰期为学期末或学期始,这时候大家对下学期的课本具有需求,我们可以发现这之间有一个很明确的关系:时间与需求的关系。即在不同的时间推荐不同的书籍[15],其流程见图 5。

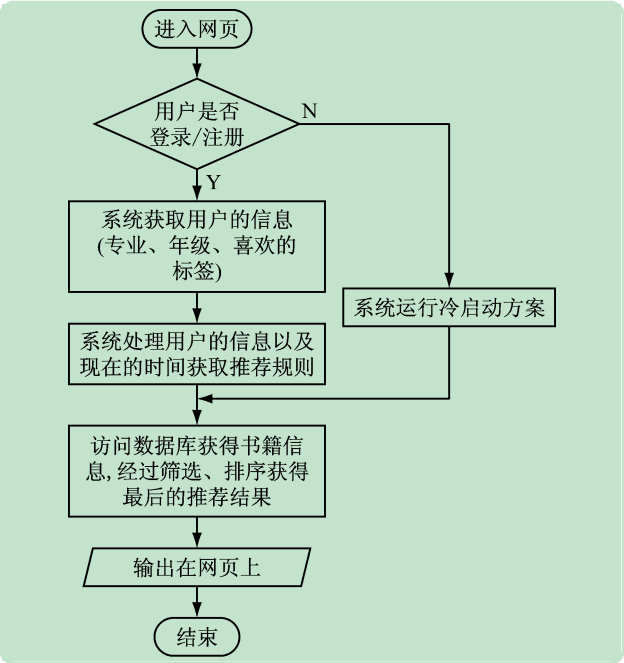

图 5　基于时间节点的推荐算法流程图

该算法将原本的根据购物数据的推荐拓展为结合时间的动态推荐,这样能在不同的时刻满足用户不同的需求,及时提供他们需要的商品推荐,或者潜在需要的商品,例如用户本学期按学校安排需要参加 CET4 的考试,那么在本学期开学前便将《星火英语 - 黑旋风试卷》(上海交大出版社出版)推荐给他,在推荐给他的同时也起到了提醒作用。

3　算法实现及检验

3.1　结合新旧因子及评分机制推荐算法的实现

为每一本书添加“用户评分”标签,使用户在购买或浏览图书时能为书籍进行评分,根据用户的评分对书籍进行排名,将用户普遍评分高的书籍推荐给有需求的用户;结合书籍热度利用加权求和的算法将推荐结果进行优化,同时结合书籍新旧因子计算推荐价值,将最终推荐价值较高的书推荐给用户,这样便可以实现对推荐结果的优化。

部分相关 php 代码如下:

```
⋮
〈head〉
〈meta　　http-equiv = " Content-Type"　content = " text/html; charset = utf-8"　/〉
〈title〉推荐规则 4_新上架的 10(非专业)本书按照评价数量(热度)来推荐〈/title〉
〈/head〉
? php
function function_recommend_rule4( db)
{
$ query = " select book_id from ( select *　from book where course_id = 111 order by register_timedesc limit 0,10)　as temp order by book_heatdesc "; //无输入,返回 rcmd_rule4 数组
$ result = mysqli_query( db,query) ;
$ num_result = mysqli_num_rows( result) ;
for( $ i = 0; $ i < $ num_result; $ i ++)
{
　　$ row = mysqli_fetch_assoc( result) ;
　　$ output_rcmd_rule4 [] = row [ 'book_id' ];
}
mysqli_free_result( result) ;
return output_rcmd_rule4;
}
? >
⋮
$ like_book = a_array_unique( like_book) ;
foreach ( like_book as key = > value) //对评分进行排序
{
```

```
        $id[$key] = $value['id'];

    $score[key] = value['score'];
    }

    array_multisort(score,SORT_DESC,SORT_NUMERIC,$like_book);
    foreach ( $like_book as $key = >value) //只返回id
        {
            $id[key] = $value['id'];
            $score[$key] = $value['score'];
        }
    return id;
    }
    ? >
    ⋮
```

3.2 打包推荐功能的实现

采用从学校教务处对课程安排进行统计，确保课程包的正确性，同时随时关注教务网的动态，一旦有课程更新，及时对课程包数据进行修正。针对推荐匹配问题，采用的解决办法为：在用户注册时要求用户注明自己的院系专业及入学年份，记录在用户数据库中，可以直接根据其记录属性进行相关推荐。

3.3 基于时间节点推荐算法的实现

首先这是根据时间节点的推荐算法，那么就要将用户的时间分为几个部分，首先是假期，在该时间段内学生处于无功课状态，此时便可以为其推荐能力提高类，旅游娱乐类，文学作品类，以再提高和休息娱乐为主要目的进行推荐，在这几类推荐给用户的同时也让他们发现新的兴趣点，得到意外惊喜发掘出潜在需求，例如该学生本来对文学作品不太感兴趣，对推荐结果中的书名《盗墓笔记》产生一点点兴趣，他开始去尝试，结果使他喜欢上了小说，从而获得惊喜。然后是在校期间，这个时间段还会继续细分成开学前、期中和期末，同时要根据安排加入社会考试时期，如四六级、计算机二级等。在开学前(期末)，就推荐课本包，将本(下)一学期要用到的课本打包推荐给用户，同时要对本(下)学期有社会考试的同学进行相关题库复习资料的推荐；在期中时则对有期中考试的同学进行题库、复习资料的推荐；期末同期中类似也是针对考试进行备考资料的推荐。

部分相关 php 代码如下：

```
    ⋮
〈head〉
    〈meta http-equiv = " Content-Type" content = " text/html; charset = utf-8" /〉
    〈title〉推荐规则3(在选课以及开学期间推送新学期所用到的专业书) < /title >
    〈/head〉
    〈body〉
    ? php
    function function_recommend_rule3( &phpmajor,& $cur_grade, $db)
    {
    $query = " select book_id from major, course, book where major. major_id = phpmajor
    andmajor. course_id = course. course_id
    andcourse. course_term = cur_grade
    andcourse. course_id = book. course_id";
    $result = mysqli_query( $db, $query);
    $num_result = mysqli_num_rows( $result);
    if( $num_result = =0)
    return;
    for( $i =0; $i < num_result; $i ++)
    {
        $row = mysqli_fetch_assoc( $result);
        $output_rcmd_rule3 [$i] = $row ['book_id'];
    }
    mysqli_free_result( result);
    return output_rcmd_rule3;
    }
    ?〉
    〈/body〉
    ⋮
```

3.4 算法检验结果分析

为了验证新设计的算法的性能，采用对比应用传统基于物品的推荐算法 Item-Based Method 以及 Item – Based Collaborative Filtering Recommendation Algorithms 这两种算法与应用设计后的算法的推荐结果指标的方法。采用本校信息学院各专业学生作为测试数据集，推荐指标采用准确 P[1] 率和平均购书时间 T_a。其中测试平均购书时间的数据集为 4 个专业随机挑选的 5 名学生，分别为：电气专业 1 名(男)，大三；通信专业 2 名(男、女)大三；自动化专业 1 名(男)，大二；电子专业 1 名(女)大一。测量准确率的数据集为随机抽取购书学生 50 名通过走访各班，以问卷调查方式得出其对推荐栏书的满意度。

测试方法为线下本地测试。在同学有较大购书需求的学期初，通过邀请本学院不同学生(数据集上文已介绍)，先使用传统推荐算法进行本地运行，使学生进行购书行为，利用后台记录其操作数据，并发放问卷对第一次满意度和准确率进行调查；再加入新的推荐算法再次使同一用户进行购书行为，记录其操作数据，发放问卷调查第二次满意度及准确率。

平均购书时间 T_a：该指标为检测新算法的加入是否提高了用户的购书效率，计算公式为

$$T_a = T/n \tag{5}$$

式中：T 为用户从登录购物网页到完成下单操作所使

用的全部时间; n 为该用户购买书籍的本书。

准确率: 推荐列表中用户喜欢的产品和所有被推荐产品的比率,计算方法为

$$P = N_{*2}/N_2 \quad (6)$$

式中: N_{rs} 推荐内容中用户喜欢的产品个数; N_r 用户喜欢的所有产品的个数; N_s 为所有被推荐产品的个数。

从图 6 中数据可以看出单纯使用基于物品的推荐算法 Item - Based Method 以及 Item - Based Collaborative Filtering Recommendation Algorithms 的算法用户平均购买时间绝大部分是要高于加入新算法后用户平均购买时间的,可以反映出用户购书效率有了一定程度的提高。

从图 7 中可以得出,准确率在加入新的算法后有了一定程度的提升,而在用户购买书本较少时由于推荐书本数一般大于用户购买数,所以导致准确率较低; 当用户购买的书本在 5 ~ 6 本时准确率最高,此时推荐满意度最好; 当购买书籍超过 6 本后准确率变化不大,维持在一个恒定水平。由于购书用户通常一次购买书籍不超过 10 本所以没有测得当用户一次购书超过 10 本的数据。同时课本打包推荐的加入也极大的提高了推荐的准确率。

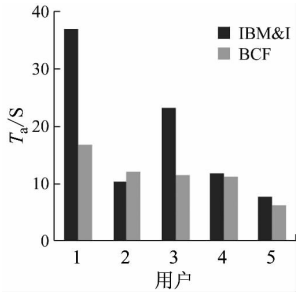

图 6　平均购书时间 T_a 前后对比

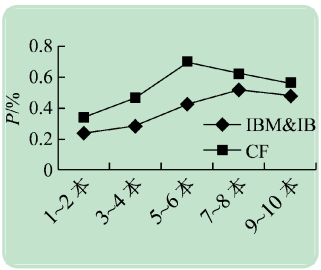

图 7　准确率随购买本书变化及前后对比图

4　结　语

针对推荐引擎直接应用普通推荐算法不能很好地满足大学校园二手书交易平台推荐需求这一问题,提出了应用结合新旧因子的推荐、基于评分机制的推荐,增加打包推荐功能以及采用基于时间节点的推荐的方法。结合新旧因子的方法将所售书籍中成色较好的优先推荐给用户; 引入评分机制后可以将大多数用户评价较高的书籍推荐给用户; 增加一键购书功能后可以使用户更方便更快捷地完成购物; 应用基于时间节点的推荐算法[16]后使推荐结果更加贴近用户群体(大学生) 的学习、生活,能及时满足用户的购书需求。通过对用户购买数据的统计,以准确率和平均购买时间为评价指标,得出的结果显示此次算法的改进进一步提高了推荐引擎的推荐质量。以上方法的使用极大地优化了二手书交易平台的推荐功能,使用户有了更好的使用体验,为大学校园二手书交易发展提供了推动力。

而在本次研究应用中,由于学生购书本书一般不超过 10 本,故所获得数据缺乏购书本书更多的数据,不能进一步验证当一次性购书较多时本系统对推荐准确率和效率的影响; 而当购书本书较少时,由于推荐本书较多,所以准确率会偏低,这也是未解决的问题之一。希望后续工作者能研究解决。

参考文献(References) :

[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] Resnick P, Iacovou N, *et al*. GroupLens: an open architecture for collaborative filtering of netnews [C] // Proceedings of the 1994 ACM con-Ference on Computer Supported Cooperative Work. ChapelHill, NorthCarolina, UnitedStates, October 22-26, 1994.

[3] Adomavicius G, Sankaranarayanan R, Shahana S, *et al*. Incorpo-rating contextual information in recommendation systems using a multidimensional approach [J]. ACM Transactions on Infor- mation Systems, 2005, 23(1) : 103-145.

[4] Weng S S, Lin B S, Chen W J. Using contextual information and multidimensional approach for recommendation [J]. ExpertSystem with Applications, 2009(36) : 1268-1279.

[5] 刘　玮. 电子商务系统中的信息推荐方法研究 [J]. 情报科学, 2006, 24(2) : 300-303.

[6] Zhao Jinghe, Liu Guiquan. Automatic modeling based on interest clustering [J]. AI Commun, 2001, 14(3) : 129-147.

[7] Zhang Yulian, Wang Quan. User profile mining of combining Web behavior and content analysis [J]. New Technology of Library and Information Service, 2007(6) : 52-55.

[8] 查大元. 个性化推荐系统的研究和实现 [J]. 计算机应用与软件, 2011, 28(1) : 47-49, 98.

[9] Zhou T C, Ma H, King I, *et al*. UserRec: A user recommendation framework in social tagging systems [C] //In: Proc. of the 24th AAAI Conf. on Artificial Intelligence. AAAI Press, 2010. 1486-1491.

[10] Wu L, Chen EH, Liu Q, *et al*. Leveraging tagging for neighborhood-aware probabilistic matrix factorization. In: Proc. of the ACM Int'l Conf. on Information and Knowledge Management. ACM Press, 2012. 1854-1858. [doi: 10. 1145/ 2396761. 2398531]

[11] 项　亮. 推荐系统实践 [M]. 北京: 人民邮电出版社, 2012.

[12] Luke Welling, Laura Thomson. PHP 和 MySQL Web 开发 [M]. 北京: 机械工业出版社, 2009.

[13] 许海玲. 互联网推荐系统比较研究 [J]. 软件学报, 2009, 20(2) : 350-362.

[14] Liu Q, Chen EH, Xiong H, *et al*. Enhancing collaborative filtering by user interests expansion via personalized ranking [J]. IEEE Trans on Systems, Man and Cybernetics—B, 2012, 42 (1) : 218-233. [doi: 10. 1109 /TSMCB. 2011. 2163711]

[15] Li B, Zhu XQ, Li RJ, *et al*. Cross-domain collaborative filtering over time [C] //In: Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence. IJCAI/AAAI Press, 2011. 2292-2298. [doi: 10. 5591 /978-1-57735-516-8 /IJCAI11-382]

[16] 孙光福, 吴　乐, 刘　淇, 等. 基于时序行为的协同过滤推荐算法 [J]. 软件学报, 2013, 24(11) : 2721-2733.