

一种融合多种用户行为的协同过滤推荐算法

高山^{1,2,3} 刘炜² 崔勇¹ 张茜⁴ 王宗敏²

(郑州大学信息工程学院 郑州 450001)¹ (郑州大学信息网络省重点开放实验室 郑州 450001)²
(河南工业大学信息科学与工程学院 郑州 450001)³ (中原工学院计算机学院 郑州 450007)⁴

摘要 协同过滤作为目前应用最成功的个性化推荐技术,在电子商务、社交网络等领域得到了广泛应用。然而,当此类算法应用到个性化医疗推荐领域时,由于个人医疗行为本身的复杂性和多样性,出现了推荐准确率下降的问题。针对这一问题,提出一种融合多种用户行为的协同过滤推荐算法,使用权重因子来综合衡量不同用户行为对推荐质量的影响,并引入重合依赖度的概念来修正传统的相似度量方法。在收集的 Top-md 数据集上的实验结果表明,该算法能够全方位表达用户的就医偏好和意愿,有效提高个性化医疗推荐系统的推荐质量。

关键词 推荐系统,协同过滤,重合依赖度,多种用户行为,权重因子

中图分类号 TP311 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.9.045

Collaborative Filtering Algorithm Integrating Multiple User Behaviors

GAO Shan^{1,2,3} LIU Wei² CUI Yong¹ ZHANG Qian⁴ WANG Zong-min²

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)¹

(Henan Provincial Key Lab on Information Networking, Zhengzhou University, Zhengzhou 450001, China)²

(College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)³

(School of Computer Science, Zhongyuan University of Technology, Zhengzhou 450007, China)⁴

Abstract Collaborative filtering is one of the most successful techniques among personalized recommender systems, and is widely used in the field of e-commerce, social networks etc. Due to the complexity and diversity of the personal health behaviors, it causes low accuracy of the recommendation algorithms when applied to personalized medicine recommendation. To deal with this problem, a new collaborative filtering algorithm integrating multiple user behaviors was proposed. The weighting factor and the overlap-dependency are introduced in classical similarity computing, and they can measure the effects of different user behaviors on the recommendation quality. Experiments on the Top-md dataset show that the new algorithm can fully express the user's preferences and wishes for medical treatment, and can effectively improve the quality of personalized medicine recommender systems.

Keywords Recommender systems, Collaborative filtering, Overlap-dependency, Multiple user behaviors, Weighting factor

1 引言

面对互联网上日益激增的信息量,以及用户对信息、商品和服务需求方式的多元化、个性化要求,个性化推荐系统(Personalized Recommender Systems)已经成为解决“信息超载”(Information Overload)问题的重要手段,并广泛应用于电子商务、社交网络、音视频点播、在线学习等领域中^[1]。与此同时,医疗数据迅猛增长,使得数据驱动个性化医疗服务成为可能。然而,用户规模和各类健康数据的不断增长,导致用户在海量医疗数据中找到自己感兴趣的信息越发困难^[2]。协同

过滤推荐算法作为目前应用最成功、最广泛的个性化推荐技术^[3],其通过分析用户对相似项目(商品、音视频、服务等)的偏爱程度,找出与目标用户偏好一致的邻居集合,然后根据这些相似用户对项目的评分来预测目标用户的喜好,从而产生推荐^[4,5]。目前,协同过滤推荐算法主要分为 3 类:基于用户的协同过滤推荐算法^[6]、基于项目的协同过滤推荐算法^[7,8]和基于模型的协同过滤推荐算法^[9],并且它们依然存在数据稀疏性问题和冷启动等问题,导致推荐质量不高。为了提高推荐质量,大量学者已做了相关的研究,GUO Gui-bing 等人^[10]提出了基于信任的奇异值分解(TrustSVD)的方法,在

到稿日期:2015-07-21 返修日期:2015-10-19 本文受教育部博士点专项科研基金(20114101110007),河南省创新人才项目(2011HASTIT003),河南省科研重点项目(13A520562),河南省高等学校重点科研项目(15A520028),河南省基础与前沿技术研究项目(152300410047)资助。

高山(1977-),女,博士生,副教授,CCF 会员,主要研究方向为数据挖掘、个性化推荐,E-mail:gaoshan_zz@126.com;刘炜(1987-),男,博士,讲师,主要研究方向为网络安全、网络性能分析、数据挖掘,E-mail:wliu@zzu.edu.cn;崔勇(1983-),男,博士生,主要研究方向为虚拟化技术、云计算,E-mail:cuiyong@zzu.edu.cn;张茜(1987-),女,博士,讲师,主要研究方向为网络流媒体、网络安全,E-mail:957613999@qq.com;王宗敏(1964-),男,博士,教授,博士生导师,主要研究方向为下一代互联网,E-mail:zmwang@zzu.edu.cn。

对未知项目进行预测评分时,同时考虑评分数据和信任数据的隐性与显性影响,这种方法能够获得较高的推荐准确率。廖寿福等人^[11]提出一种基于项目属性填充的个性化推荐冷启动算法,将用户对项目的评分转换为对项目属性的评分;该方法能够提高个性化推荐系统的推荐精度,但局限于邻居数量较多的情况。

虽然基于协同过滤技术的传统推荐方法已经在 Amazon、Netflix、淘宝、豆瓣等大型网站取得了巨大的成功^[12],但由于个人医疗行为本身的复杂性和多样性,导致高维评分矩阵的数据稀疏程度日益加剧^[13,14],冷启动问题依然存在,使得传统个性化推荐方法在医疗推荐应用场景下的推荐质量不高,大大降低了推荐系统的有效性。本文将协同过滤技术应用到个性化医疗推荐领域中,为了进一步缓解数据稀疏性问题和冷启动问题,在基于项目的协同过滤算法(Item-based Collaborative Filtering, IBCF)的基础上,引入重合依赖度(Overlap-Dependency)的概念对传统的相似度计算方法进行了改进;并且在计算项目间相似度时,同时考虑两类不同的用户就医行为,提出了一种融合多种用户行为的协同过滤推荐算法(Collaborative Filtering Algorithm Integrating Multiple User Behaviors, MUBCF)。该算法将不同类型的用户就医行为进行互补,在一定程度上可以缓解在某一行为模式下由于用户对现有项目评分数据稀少导致的冷启动问题。通过实验室自行设计开发的健康服务平台,收集了用户多种就医行为下的数据。在该数据集上的实验结果表明,本文提出的算法的评分预测准确度较高,进一步提高了推荐质量,适用于个性化医疗推荐应用领域。

2 传统相似度计算方法的改进

2.1 传统相似度计算存在的问题

Sarwar 等人^[7]在 2001 年提出了基于项目的协同过滤算法(Item-based Collaborative Filtering, IBCF),由于项目间的相似性相对稳定,并可离线完成计算,尤其适用于用户数远多于项目数的应用场景。IBCF 算法具有的特点和优势更加适用于个性化医疗推荐领域,因此本文基于 IBCF 算法进行相似度计算方法的研究。在基于项目的协同过滤算法 IBCF 中,项目间相似度的传统计算方法主要有:余弦相似度(COS)、修正余弦相似度(adjCOS)和 Pearson 相关系数(PCC)等^[7,15]。其中余弦相似度的取值范围为 $[0, 1]$,值越大表示相似性越大。Pearson 相关系数的取值范围为 $[-1, 1]$,值为 0 表示不相关,值为负表示两向量负相关,值为正表示两向量正相关。修正余弦相似度是在余弦相似度方法的基础上改善了不同用户评分尺度的差异性问题。已有文献使用 MovieLens 数据集,通过实验衡量了 3 种不同的相似度计算方法在 IBCF 协同过滤算法中预测的准确性^[7]。

传统的相似度计算方式都只是在同时访问两个项目的公共用户集合范围内衡量不同项目间的相似性,若公共用户数偏少,则由此计算得到的相似性不能全面和准确地度量二者的相关性。针对这一问题,应通过一定策略对相似性进行必要的修正,以反映相似性的可信度。国内外的研究学者曾分别采用固定阈值^[16]或动态阈值调整的方法^[9]来区分不同公

共用户数量情况下相似度的可信度,但是固定阈值的解决方法无法适应不同用户行为模式下项目间公共用户数量的差异性,而动态阈值调整的方法需要算法进行多次实验才能获取。

针对以上问题,本文从个性化医疗推荐应用场景的特点出发,在研究改进由于公共用户数量偏少导致的相似性计算不准确的问题时,一方面,从两个项目间的所有评分用户角度,衡量公共用户数量在两个项目间所有评分用户中所占的比重;另一方面,衡量两项目间的公共用户数量在全体项目的共同评分用户数中所占的比重,引入重合依赖度(Overlap-Dependency)的概念对传统的相似度计算方法进行修正,以区分项目间相似度计算结果的可信度。

2.2 相关定义

定义 1(重合度) 设 i 和 j 为项目集合 I 中的任意两个项目,对两个项目评过分的用户集合分别为 $U_i = \{u_c \mid u_c \in U \wedge r_{c,i} \neq 0\}$ 和 $U_j = \{u_c \mid u_c \in U \wedge r_{c,j} \neq 0\}$,则项目 i 和 j 之间的公共用户评分重合度为:

$$o(i, j) = \frac{2|U_i \cap U_j|}{|U_i| + |U_j|}, 0 \leq o(i, j) \leq 1 \quad (1)$$

定义 2(全局依赖度) 设 i 和 j 为项目集合 I 中的任意两个项目, $|\bar{U}_{ALL}|$ 为项目 i 和项目 j 间共同评分用户数的平均值,则项目 i 和项目 j 之间的公共用户评分全局依赖度为:

$$d(i, j) = \begin{cases} \frac{|U_i \cap U_j|}{|\bar{U}_{ALL}|}, & \text{if } |U_i \cap U_j| < |\bar{U}_{ALL}| \\ 1, & \text{if } |U_i \cap U_j| \geq |\bar{U}_{ALL}| \end{cases}, \quad 0 \leq d(i, j) \leq 1 \quad (2)$$

定义 3(重合依赖度) 设 i 和 j 为项目集合 I 中的任意两个项目,则项目 i 和项目 j 之间的公共用户评分重合依赖度为:

$$\begin{aligned} od(i, j) &= o(i, j) \times d(i, j) \\ &= \begin{cases} \frac{2|U_i \cap U_j|}{(|U_i| + |U_j|)} \times \frac{|U_i \cap U_j|}{|\bar{U}_{ALL}|}, & \text{if } |U_i \cap U_j| < |\bar{U}_{ALL}| \\ \frac{2|U_i \cap U_j|}{(|U_i| + |U_j|)}, & \text{if } |U_i \cap U_j| \geq |\bar{U}_{ALL}| \end{cases}, \\ &0 \leq od(i, j) \leq 1 \end{aligned} \quad (3)$$

2.3 使用重合依赖度的概念修正传统的相似度计算方式

引入重合依赖度的概念之后,可以对余弦相似度、修正余弦相似度和 Pearson 相关系数 3 种传统的相似度度量方法进行修正,计算公式分别为:

$$COSSim'(i, j) = od(i, j) \times \frac{\sum_{u \in U_{i,j}} r_{u,i} r_{u,j}}{\sqrt{\sum_{u \in U_{i,j}} r_{u,i}^2} \sqrt{\sum_{u \in U_{i,j}} r_{u,j}^2}} \quad (4)$$

$$adjCOSSim'(i, j) = od(i, j) \times \frac{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_u)(r_{u,j} - \bar{r}_u)}{\sqrt{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{u \in U_{i,j}} (r_{u,j} - \bar{r}_u)^2}} \quad (5)$$

$$PCCSim'(i, j) = od(i, j) \times \frac{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_i)(r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{i,j}} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in U_{i,j}} (r_{u,j} - \bar{r}_j)^2}} \quad (6)$$

其中, $U_{i,j}$ 表示同时对项目 i 和项目 j 评分的用户集合。

通过式(1)计算获得的重合度不仅正比于两项目间的公共评分数量,还反比于两项目各自的用户评分数量,使得评分重合度可以适应不同项目的评分数量分布。引入重合度的概念后,在衡量项目之间的相似性时,两个项目间的公共评分用户数在所有对这两个项目评过分的用户数中所占的比例越高,则计算得到的相似度越能反应项目间的真实相关性。

引入全局依赖度的概念后,在衡量项目之间的相似性时,两项目间的公共用户数量在全体项目的共同评分用户数中所占的比重越高,越能说明这两个项目受用户偏爱的程度越高,则认为项目之间的相似性也越高。在个性化医疗推荐应用领域中,该参数可以表示患者对医生的依赖程度。此外,全局依赖度的概念将所有项目间公共用户数的平均值作为参考阈值,当计算项目间相似性时,如果公共用户数量在平均值以下,则说明此时数量偏少的公共用户评分不足以全面和准确地衡量其相关性,全局依赖度此时相当于一个惩罚系数,根据公共用户数量所占的比重,进一步区分项目间相似性的可信度。

综上所述,在度量两个项目间的相似性时,同时考虑重合度和全局依赖度,即引入重合依赖度的概念,既能强调公共用户评分重合度的重要性,又能反映两个项目在所有用户间的整体流行度和受依赖程度。通过实验结果可以看出,改进后的相似度计算方法较传统的度量方法能够获得更高的评分预测精度,进一步提升推荐质量。

3 引入重合依赖度的 MUBCF 算法

3.1 MUBCF 算法

协同过滤算法进行推荐的主要依据是用户对项目的评分,但在单一行为模式下的用户评分数据的稀疏性往往比较高,通常商用推荐系统的稀疏度可以达到 99% 以上,这种极度稀疏的评分矩阵将给协同过滤推荐造成很大的影响^[14]。

在个性化医疗推荐应用场景中,用户的就医行为可以有很多种,本文选取了其中最具有代表性的两类线上就医行为:在线咨询和预约挂号。利用不同行为模式下的用户评分数据,分别计算项目间的相似度,并对两部分相似度进行线性组合,将用户不同类型的就医行为进行互补,从而进一步缓解用户在某一种行为下评分数据稀少带来的冷启动问题对评分预测准确性的影响。

本文以基于项目的协同过滤算法 IBCF 为基础,提出一种适用于个性化医疗推荐应用领域的融合多种用户行为的协同过滤推荐算法(MUBCF),并使用重合依赖度来修正传统的相似度计算方法,根据用户多种就医行为下的历史偏好信息,为其推荐最感兴趣的 Top-N 医生列表。

融合多种用户行为的个性化医疗推荐问题的形式化描述如下。

已知:2 种不同用户行为下产生的预约挂号评分数据 $(u, i, r_{u,i})$ 和在线咨询评分数据 $(u, i, c_{u,i})$ 。

求:目标用户 u 的 Top-N 医生推荐集。

其中,用 U, I 分别表示所有的用户和医生, R (Registration) 表示用户在预约挂号行为模式下产生的评分矩阵, C (Consulting) 表示用户在在线咨询行为模式下产生的评分矩阵。

本文使用权重因子 θ 将两类不同就医行为模式下的项目间相似性进行线性组合,则 MUBCF 算法中相似度度量方法的计算公式为:

$$\omega(i, j) = \theta S(i, j) + (1 - \theta) D(i, j) \quad (7)$$

其中, $\omega(i, j)$ 表示融合两类不同用户行为之后的项目间相似性, $S(i, j)$ 表示预约挂号行为模式下的项目间相似性, $D(i, j)$ 表示在线咨询行为模式下的项目间相似性, $S(i, j)$ 和 $D(i, j)$ 可以分别使用重合依赖度修正的传统相似度度量方法进行计算。权重因子 θ 的取值范围为 $[0, 1]$, 当 $\theta = 1$ 时, MUBCF 算法转变为单一挂号行为模式下的协同过滤算法; 当 $\theta = 0$ 时, MUBCF 算法转变为单一咨询行为模式下的协同过滤算法。本文利用权重因子 θ 将用户不同类型的医疗行为进行互补和融合, 动态衡量两类不同的用户行为对个性化医疗推荐质量的影响, 最佳权重值需要经过多次实验来获取。

实验结果表明, 在计算项目间的相似度时, 融合多种用户就医行为不仅能够进一步缓解用户在单一行为模式下评分数据稀少带来的冷启动问题, 使得项目间的相似性计算更加精确, 而且能够全方位表达用户的就医偏好和意愿, 有效提高了推荐质量。

3.2 算法描述

综合前面的论述, 算法 1 给出了为目标用户 u 产生 Top-N 医生推荐集的过程, 在个性化医疗推荐的应用场景下, 项目代表的是医生资源。

算法 1 MUBCF 算法

输入: 目标用户 u 、挂号评分矩阵 R 、咨询评分矩阵 C

输出: 用户 u 的 Top-N 医生推荐集

Step1: 根据用户-项目挂号评分矩阵 R 和用户-项目的咨询评分矩阵 C , 对于项目集合 I 中的所有项目, 计算任意两个项目之间在 2 种不同行为模式下的共同评分用户数的平均值 $|\bar{U}_{ALL}|$ 。

Step2: 根据本文提出的新的相似度计算方法, 以及挂号评分矩阵 R 和咨询评分矩阵 C , 使用式(4)一式(7)计算任意两个项目间的相似度, 得到相似度矩阵 S ;

Step3: 在评分矩阵 R 中获取目标用户 u 的已评分项目 I_u , 对所有项目 $i \in I_u$, 读取相似度矩阵 S 得到该项目的 K 最近邻居集合 $N_i = \{i_1, i_2, \dots, i_k\}$, 合并所有 N_i , 并从中删除 I_u 中已经存在的项目, 得到候选项目集 C_a ;

Step4: 对所有项目 $j \in C_a$, 在 I_u 中为该项目选择相似近邻, 然后根据式(8)以及相似度矩阵 S , 预测目标用户 u 对项目 i 的评分;

$$P_{u,i} = \bar{r}_i + \frac{\sum_{j \in N} \omega_{i,j} \times (r_{u,j} - \bar{r}_j)}{\sum_{j \in N} |\omega_{i,j}|} \quad (8)$$

其中, N 表示用户已评分项目集中待预测项目的近邻集。

Step5: 将候选项目集中 C_a 中所有项目按预测评分从大到小排序, 将其中最前面的 n 个项目作为目标用户 u 的 Top-N 推荐集。

其实现流程如图 1 所示。

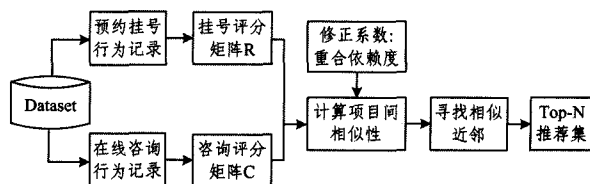


图 1 MUBCF 算法的实现流程

4 实验结果及分析

4.1 数据集说明

作者所在实验室自行设计开发了一个医疗健康类服务平台,截止到2014年12月,平台已联合了郑州大学6家附属医院的优质医疗资源,坐诊医生数2288人,注册用户数38490人,预约挂号数为43714次,在线咨询数为16700次。

基于该平台的真实历史数据,本文主要关注预约挂号和在线咨询两种最具代表性的用户线上就医行为,分别抽取了其中5000名注册用户对于500名医生的预约挂号评分数据和在线咨询评分数据,该数据能够描述真实个性化医疗推荐的应用场景,本文将该数据集命名为Top-md。该数据集中的评分值为1至5的整数,数值越高表示用户对医生的偏爱程度越高。实验中将数据集按照80%和20%的比例划分为训练集和测试集,利用5折交叉法进行了实验。表1给出了Top-md数据集的基本特征。

表1 Top-md数据集的基本特征

	用户数	项目数	评分数	稀疏度
预约挂号	5000	500	32711	0.9869
在线咨询	5000	500	13260	0.9947

4.2 评价标准

实验采用推荐系统中常用于评估推荐质量的平均绝对误差(Mean Absolute Error, MAE)作为度量标准^[17]。它通过计算推荐系统给出的预测评分与用户的真实评分之间的绝对距离的均值来度量推荐的准确性,MAE与推荐的预测准确性成反比,MAE的值越小,推荐质量越高。

假设对于测试集中的一个用户 u 和项目 i , $r_{u,i}$ 是用户 u 对项目 i 的真实评分, $\hat{r}_{u,i}$ 是推荐算法给出的预测评分,则平均绝对误差MAE的计算公式为:

$$MAE = \frac{\sum_{i=1}^N |r_{u,i} - \hat{r}_{u,i}|}{T} \quad (9)$$

其中, T 为测试集中的评分记录个数。

4.3 实验结果

4.3.1 3种传统相似度计算方法的比较

本文首先对余弦相似度(COS)、修正余弦相似度(adj-COS)和Pearson相关系数(PCC)3种传统相似度度量方法的预测准确度进行比较,目的是选取在Top-md数据集上具有最佳推荐质量的相似度度量方法进行改进。横坐标为最近邻个数 K ,其取值范围从5递增至50,间隔为5,纵坐标为MAE的值,实验结果如图2所示。

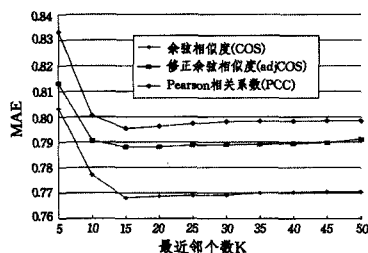


图2 3种传统相似度计算方法的比较

如图2所示,在3种度量方法中,使用余弦相似度(COS)

的相似度度量方法计算得到的MAE值最小,表示其评分预测的准确率最高,在推荐质量上要优于其他两种相似度计算方法。所以,本实验以余弦相似度计算方法为基础进行进一步的研究。

4.3.2 余弦相似度度量方法改进后的效果

本文引入重合依赖度的概念对传统的相似度计算方法进行修正,在图2所示实验结果的基础上,将余弦相似度和改进后的余弦相似度两种度量方法的评分预测准确率进行对比,最近邻个数 K 的取值范围从5递增至50,纵坐标为MAE的值,实验结果如图3所示。

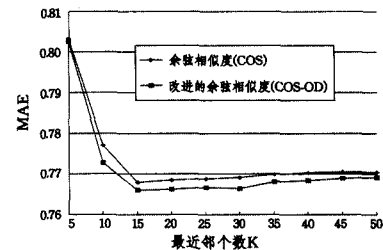


图3 改进前后的余弦相似度的比较

如图3所示,改进后的相似度度量方法计算得到的MAE值有所下降,表示使用重合依赖度对传统的余弦相似度度量办法进行修正之后,其评分预测的准确率有所提升,能够有效提高推荐系统的推荐质量。

4.3.3 获取权重因子 θ 的最优值

本文提出的融合多种用户行为的协同过滤推荐算法MUBCF使用权重因子 θ 来动态衡量两类不同的用户行为对推荐质量的影响。权重因子 θ 的取值范围从0.1递增至1,步长为0.1,最近邻个数 K 取15,因为在此取值下推荐算法的MAE值最低。将不同权重因子取值对评分预测准确率的影响效果进行对比,实验结果如图4所示。

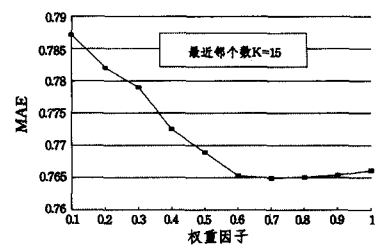


图4 权重因子 θ 值的选取(最近邻个数 $K=15$)

如图4所示,当权重因子 θ 值由0.1递增至0.7时,推荐算法评分预测的MAE值逐步降低,之后随着 θ 值的增长,MAE值的变化不大。MAE值最小时,权重因子 θ 值为0.7,说明在该取值下融合两种用户行为的方法计算得到的项目间的相似性更加精确,从而能够获得更高的评分预测准确率。

4.3.4 与其他算法的比较

为了验证本文提出的MUBCF算法的有效性,分别将其与IBCF协同过滤算法和使用重合依赖度改进的IBCF协同过滤算法进行了对比,实验结果如图5所示。从图5可以看出,本文提出的融合多种用户行为的协同过滤推荐算法MUBCF相对于传统的IBCF协同过滤算法以及仅使用重合依赖度进行改进的IBCF算法,其算法的推荐质量有明显的提高,更适于用个性化医疗推荐应用领域。

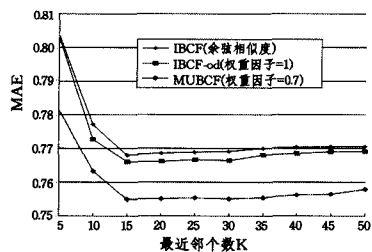


图5 MUBCF推荐算法的评分预测准确度对比

结束语 本文在传统相似度计算方法与构成的基础上, 基于 IBCF 协同过滤算法, 提出一种融合多种用户行为的协同过滤推荐算法。该算法首选结合个性化医疗推荐的特殊应用背景, 引入重合依赖度的概念对传统的相似度度量方法进行改进; 同时, 基于个性化医疗行为的多样性, 将用户多种行为进行互补和融合, 综合衡量其对推荐质量的影响。在收集的数据集 Top-md 上的实验结果表明, 该算法能够进一步缓解用户在单一行为模式下对已有项目评分数据稀少带来的冷启动问题, 使得改进后项目间的相似性计算更加精确, 从而可以全方位表达用户的就医偏好和意愿, 提高个性化医疗推荐系统的推荐质量。

参考文献

[1] Wang Guo-xia, Liu He-ping. Survey of Personalized Recommendation System [J]. Computer Engineering and Applications, 2012, 48(7): 66-76 (in Chinese)
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, 48(7): 66-76

[2] Huang Xin-ting, Bao Xiao-yuan, Yu Guo-pei, et al. Research on Personalized Medical Service Engine Driven by Medical Big Data [J]. China Digital Medicine, 2014, 9(8): 5-7 (in Chinese)
黄新霆, 包小源, 俞国培, 等. 医疗大数据驱动的个性化医疗服务引擎研究[J]. 中国数字医学, 2014, 9(8): 5-7

[3] Bobadilla J, Ortega F, Hernando A. A collaborative filtering similarity measure based on singularities [J]. Information Processing and Management, 2012, 48(2): 204-217

[4] Xu Hai-ling, Wu Xiao, Li Xiao-dong, et al. Comparison Study of Internet Recommendation System [J]. Journal of Software, 2009, 20(2): 350-362 (in Chinese)
许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362

[5] Huang Chuang-guang, Yin Jian, Wang Jing, et al. Uncertain Neighbor' Collaborative Filtering Recommendation Algorithm [J]. Chinese Journal of Computers, 2010, 33(8): 1369-1377 (in Chinese)
黄创光, 印鉴, 汪静, 等. 不确定近邻的协同过滤推荐算法[J]. 计算机学报, 2010, 33(8): 1369-1377

[6] Zhao Z D, Shang M S. User-based Collaborative - Filtering Recommendation Algorithms on Hadoop [C] // Third International Conference on IEEE Knowledge Discovery and Data Mining

(WKDD'10). 2010: 478-481

[7] Sarwar B, Karypis G, Konstan J, et al. Item-based Collaborative Filtering Recommendation Algorithms [C] // Proceedings of the 10th International World Wide Web Conference. Paris: IEEE Computer Society Press, 2001: 285-295

[8] Deng Ai-lin, Zhu Yang-yong, Shi Bai-le. A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction [J]. Journal of Software, 2003, 14(9): 1621-1628 (in Chinese)
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): 1621-1628

[9] Li Peng-fei, Wu Wei-min. Optimized Implementation of Hybrid Recommendation Algorithm [J]. Computer Science, 2014, 41(2): 68-71 (in Chinese)
李鹏飞, 吴为民. 基于混合模型推荐算法的优化[J]. 计算机科学, 2014, 41(2): 68-71

[10] Guo Gui-bing, Zhang Jie, Yorke-Smith N. TrustSVD: Collaborative Filtering with Both the Explicit and Implicit Influence of User Trust and of Item Ratings [C] // Proceedings of Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015: 123-129

[11] Liao Shou-fu, Lin Shi-ping, Guo Kun. A Cold-Starting Personalized Recommendation Algorithm [J]. Journal of Chinese Computer Systems, 2015, 36(8): 1723-1727 (in Chinese)
廖寿福, 林世平, 郭昆. 个性化推荐冷启动算法[J]. 小型微型计算机系统, 2015, 36(8): 1723-1727

[12] Herlocker J, Konstan J, Borchers A, et al. An Algorithmic Framework for Performing Collaborative Filtering [C] // Proceedings of the 22nd annual international ACM SIGIR Conference on Research and Development in Information Retrieval. 1999: 230-237

[13] Liu Jian-guo, Zhou Tao, Wang Bing-hong. Research Progress in Personalized Recommendation System [J]. Progress in Natural Science, 2009, 19(1): 1-15 (in Chinese)
刘建国, 周涛, 汪秉宏. 个性化推荐系统的研究进展[J]. 自然科学进展, 2009, 19(1): 1-15

[14] Guo G. Integrating Trust and Similarity to Ameliorate the Data Sparsity and Cold Start for Recommender Systems [C] // Proceedings of the 7th ACM Conference on Recommender Systems. 2013: 451-454

[15] Breese J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering [C] // Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI'98). 1998: 43-52

[16] Herlocker J, Konstan J, Riedl J. An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms [J]. Information Retrieval, 2002, 5(4): 287-310

[17] Sarwar B, Karypis G, Konstan J, et al. Application of dimensionality reduction in recommender system - A case study [C] // ACM WebKDD 2000 Workshop. 2000