一种基于空间变换的协同过滤推荐算法

赵兴旺 梁吉业 郭兰杰
(山西大学计算机与信息技术学院 太原 030006)
(计算智能与中文信息处理教育部重点实验室(山西大学) 太原 030006)

摘 要 传统的协同过滤推荐算法在实际应用中往往面临着计算可扩展性的问题。为解决此问题,文中在基于物品的协同过滤推荐的框架下,通过融合社交关系信息,提出了一种基于空间变换的协同过滤推荐算法。首先,根据用户社交网络信息,运用社区发现算法将用户划分为不同的类;其次,基于评分信息,根据用户和物品之间的对应关系找到各个用户类所对应的物品类;最后,通过各个物品对每一物品类的隶属关系,将稀疏的高维评分信息矩阵转换为一个低维稠密的物品隶属度矩阵,进而基于该矩阵进行相似度计算并进行协同过滤推荐。在公开数据集上将所提方法与其他算法进行了对比实验分析,结果表明,所提算法能够在保证推荐准确性的同时明显提升计算效率。

关键词 协同过滤,社交网络,空间变换,可扩展性
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.07.003

Collaborative Filtering Recommendation Algorithm Based on Space Transformation

ZHAO Xing-wang LIANG Ji-ye GUO Lan-jie
(School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)
(Key Laboratory of Computational Intelligence and Chinese Information Processing (Shanxi University), Ministry of Education, Taiyuan 030006, China)

Abstract In real applications, traditional collaborative filtering recommendation algorithms are usually faced with the problem of computational scalability. To solve this problem, in the framework of item-based collaborative filtering recommendation, a collaborative filtering recommendation algorithm based on space transformation was proposed in this paper. Concretely speaking, according to the user social network information, the users are firstly divided into different clusters by using the community discovery algorithm. Then, item clusters are found according to the corresponding relationship between users and items in the rating information matrix. And the membership of each item for each item clusters is calculated. The sparse high dimensional rating information matrix is transformed into a low dimensional dense membership matrix, and then the similarities between items are carried on the transformed matrix. The proposed algorithm was compared with other algorithms on the public data set. The experimental results show that the proposed algorithm can significantly improve the computational efficiency while guaranteeing the accuracy of recommendation.

Keywords Collaborative filtering, Social network, Space transformation, Scalability

1 引言

随着移动互联网、云计算、大数据等一系列新兴技术的快速发展与广泛应用,推荐系统已经渗透到人们日常生活的各方面。目前,推荐系统已经在电子商务、娱乐网站、在线广告、社交网站、新闻网站、服务推荐等领域得到了广泛应用,不仅给运营商带来了商业利益,也给用户带来了诸多便利和个性化体验[1-3]。

近年来,研究者们根据不同的需求提出了一系列推荐算法,主要包括协同过滤推荐算法、基于内容的推荐算法以及混合推荐算法等。其中,协同过滤推荐算法具有易实现、跨领域等诸多优势,现已成为发展最快、应用最广的一类推荐算法[4]。随着互联网规模的快速扩大,协同过滤推荐算法在实际应用中面临着数据稀疏、可扩展性等问题[3,5]。

为解决以上问题,研究者们已开展了一些探索性工作。其中,为了缓解协同过滤推荐中的稀疏性,冷亚军等人[6]提出了两阶段最近邻选择算法,其首先找到用户近邻倾向性高的集合,然后计算它们之间的等价关系,从而得到最终的最近邻

到稿日期:2017-07-16 返修日期:2017-09-22 本文受国家自然科学基金项目(61432011,U1435212,61603230),山西省自然科学基金项目(201601D202039),山西省教育厅高校科技创新项目(2016111),山西省研究生教育创新项目(2018BY007)资助。
赵兴旺(1984-),男,博士生,讲师,CCF会员,主要研究方向为数据挖掘与机器学习,E-mail:zhaoxw84@163.com;梁吉业(1962-),男,博士,教授,CCF会员,主要研究方向为粒计算、数据挖掘与机器学习,E-mail:ljy@sxu.edu.cn(通信作者);郭兰杰(1991-),男,硕士,主要研究方向为社会化推荐,E-mail:guolanjiesxu@163.com。

集合，有效提高了近邻搜寻的准确性。Wang 等人[7]通过相似用户对相似物品的评分进行预测，弥补了单一协同过滤方法中邻居数量不足的缺陷，在一定程度上缓解了稀疏性问题。Liang 等人[8]利用联合聚类方法对原始评分矩阵进行聚类，并将类别相似性和传统评分相似性进行融合，有效缓解了传统相似性计算不准确的问题。Koren 等人[9]将原始稀疏评分矩阵分解为低维稠密的潜因子矩阵，解决了协同过滤算法对数据稀疏性敏感的问题。针对计算可扩展性问题，Zeng 等人[10]认为推荐系统中用户的贡献度是有区别的，将贡献度最大的 20％的用户组成核心用户群，就可以达到利用全部用户 90％的推荐精度。Cai 等人[11]采用降维的思想，利用物品的类别属性信息和评分信息将用户映射为对应类别的用户群，从而将用户特征向量从高维空间转化到低维空间，大大减少了算法的执行时间。Xu 等人[12]利用聚类分析技术将用户和物品分别划分为一些相似的类，并将传统协同过滤算法中的相关计算转变到类内部进行，减小了计算的规模，从而缩短了算法的运行时间，有效提高了推荐结果的精度。

但是，以上改进的协同过滤推荐算法仅仅利用了单一的用户-物品评分信息，在信息源方面存在一定的局限性。近年来，社会化媒体和社交网络的快速兴起不仅丰富了人们与家人、朋友的关系，更为社会化关系的深层次分析和挖掘提供了重要的数据资源。在社交关系中，用户更可能和与其有相似偏好的用户建立社交关系，社会影响显示，有社交关系的用户之间更可能具有相似的兴趣爱好。与真实世界相同，在社交网络中，用户在做出一些决定之前往往会参考其好友的建议，因此，在协同过滤推荐中可以融合社交网络的信息来提高推荐性能。例如，Liu 等人[13]利用用户社交关系信息来改进传统协同过滤算法寻找最近邻的过程，将评分相似的用户和朋友共同作为用户邻居，缓解了数据稀疏引起的邻居数量不足的问题。Guo 等人[14]在传统的协同过滤推荐算法中结合社交网络中的用户信任关系，利用信任用户对各物品的评分来补充并代表目标用户对各物品的喜好，缓解了数据的稀疏性问题和冷启动问题。郭兰杰等人[15]在协同过滤推荐中，利用社交关系信息在物品相似度计算和评分预测阶段分别对评分矩阵中的缺失值进行选择性填充，使得评分矩阵中的已有信息得到最大化利用，有效缓解了数据稀疏性问题。为解决计算可扩展性问题，本文融合社交关系信息提出了一种基于空间变换的协同过滤推荐算法。首先，根据用户社交网络信息，运用社区发现算法将用户划分为不同的类；其次，基于评分信息，根据用户和物品之间的对应关系找到各个用户类所对应的物品类；最后，通过各个物品对每一物品类的隶属关系，将稀疏的高维评分信息矩阵变换为一个低维稠密的物品隶属度矩阵，进而基于该矩阵进行物品间相似度的计算并进行推荐。在公开数据集上将所提算法与其他算法进行了实验对比分析，结果表明，提出的算法能够在保证推荐准确性的同时明显提升计算效率。

2　相关工作

传统的协同过滤推荐主要包括用户和物品两种对象。物品可以是电影、音乐、商品等现实存在的所有资源。用户与物品之间的观看、收听、购买、浏览、收藏等交互行为通常由评分来反映。传统的协同过滤技术通常分为基于模型的方法和基于内存的方法两类。前者基于聚类分析、回归分析、潜在语义空间和贝叶斯模型等机器学习方法进行建模；后者则通过采用全部用户-物品评分矩阵来进行计算，又可以具体分为基于用户的协同过滤推荐和基于物品的协同过滤推荐两种[1-3]。本文主要在基于物品的协同过滤推荐框架下进行分析。下面对此类方法进行简要介绍，该方法主要包括 3 步。

(1)输入数据

用户和物品的交互行为通常用一个评分矩阵 $R_{m\times n}$ 来表示，$U=\{U_1,U_2,\cdots,U_m\}$ 表示用户组成的集合，$I=\{I_1,I_2,\cdots,I_m\}$ 表示物品组成的集合，矩阵中的元素 $r_{u,i}$ 表示第 u 个用户对第 i 个物品的评分，通常取值 1～5，空值则代表没有评分。在实际应用中，往往有成千上万个用户和物品，而每个用户关注的物品类型和个数都是有限的，一个用户评分过的物品只是整体的一小部分。评分数据的稀疏度大都达到了 99％以上。

(2)确定目标物品的最近邻

首先，需要利用皮尔逊相似度和余弦相似度等度量方法来计算物品间的相似度；然后按照相似度从大到小的顺序对物品排序，将最前面的 k 个作为目标物品最近邻。

(3)计算评分预测值

找到目标物品的最近邻后，就可以根据目标用户对它们的评分值来计算对目标物品的预测值。在计算得到每个未知物品的预测值后，选择最大的前 L 个物品推送给用户即可。常用的预测公式的定义如下：

$$P_{u,i}=\bar{r}_i+\frac{\sum\limits_{j\in S(i)} sim(i,j)\cdot(r_{u,j}-\bar{r}_j)}{\sum\limits_{j\in S(i)} sim(i,j)} \tag{1}$$

其中，$S(i)$ 表示物品 I_i 的 k 近邻集合，$sim(i,j)$ 表示物品 I_i 和物品 I_j 的相似度，$r_{u,i}$ 表示第 u 个用户对第 i 个物品的评分，$\bar{r}_i$ 和 $\bar{r}_j$ 分别表示物品 I_i 和物品 I_j 的所有评分的均值。

3　基于空间变换的协同过滤推荐算法

3.1　用户类及物品类的生成

社交网络中，用户之间往往存在着社区结构，社区内的用户之间联系紧密，社区之间的联系稀疏。一个社区往往由一些具有相同爱好的用户组成，如篮球爱好者社区、影视爱好者社区、计算机爱好者社区等。每个社区会有相对应的一些物品类别，例如，篮球社区会更关注运动类器材、服装等，而影视社区会更关注电影、影视玩偶、演员等。当然，不同社区的用户也可能会关注相同的物品，例如，篮球社区中也会有喜欢看电影的用户，但其权重通常会小于运动类物品。因此，根据社区内用户对物品的评分信息，能够得到该社区用户所关注的物品组，这些物品很大程度上与该社区的兴趣特征一致，即不同物品组具有不同的潜在类别特征。同时，不同物品组之间是有重叠的，这些物品往往比较大众，没有明显的类别特征。

已有研究表明，用户之间即使不存在显式的社交关系，也仍然可以构造隐式的社交网络信息，因为评分相似度越大的用户间存在社交关系的机率会越大。因此，为了使社区划分

的结果更加准确，对于原始社交网络中没有显式的社交关系但评分相似度较大的用户对，本文在他们之间人为地加入隐式社交关系，使得社交网络中不仅有用户自己的朋友，还包括一些兴趣相同但是由于种种原因还没有结识的“朋友”，以期改善用户的社区划分结果，进而改善物品组的划分。

为了对物品进行划分，首先需要对用户的社交网络进行社区划分，本文采用了 Smart Local Moving (SLM)算法[16]。该算法通过最大化模块度来完成，能够快速地进行社区发现，现已成功应用到有百万级节点和亿万级边的网络中。根据社区发现算法，本文中的用户类表示为 $C=\{c_1,c_2,\cdots,c_k\}$，其中 k 为用户类(社区)个数，且 $c_1\cup c_2\cup\cdots\cup c_k=C$。用户类 $c_i=\{u_1^i,u_2^i,\cdots,u_{p^i}^i\}$，其中 u_1^i 表示用户类 c_i 中的第 1 个用户，p^i 表示第 i 个用户类中的用户个数。物品类表示为 $G=\{g_1,g_2,\cdots,g_k\}$，其中 k 为物品类个数，与用户类个数相同，且它们是可重叠的。假设 RI_1^i 表示用户 u_1^i 已评分的物品集合，则第 i 个物品类可表示为 $g_i=RI_1^i\cup RI_2^i\cup\cdots\cup RI_p^i$。用户类与物品类之间的对应关系如图 1 所示。

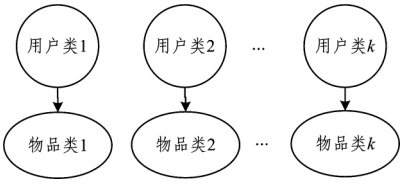

图 1 用户类与物品类的对应关系

Fig. 1 Relationship of user clusters and item clusters

3.2 物品隶属度矩阵的构建

根据评分信息可以计算各个物品对各个物品类的隶属程度。本文基于两个特征来构建物品隶属度矩阵。第一个特征表示第 i 个物品 I_i 被第 j 个物品类 g_j 所对应的用户类 c_j 中用户对其的评分数量占 I_i 所有评分的比例；第二个特征表示第 i 个物品 I_i 被第 j 个物品类 g_j 所对应的用户类 c_j 中用户对其评分的平均值，并使用评分矩阵中的最大评分值进行归一化。定义第 i 个物品 I_i 对第 j 个物品类 g_j 的隶属度 $w_{i,j}$ 为：

$$w_{i,j}=\frac{1}{2}\times\frac{n_{j,i}}{n_i}+\frac{1}{2}\times\frac{\bar{r}_{j,i}}{r_{\max}} \tag{2}$$

其中，$n_{j,i}$ 表示第 j 个物品类 g_j 所对应的社区 c_j 中用户对第 i 个物品 I_i 的评分数量，n_i 表示第 i 个物品 I_i 总的评分数量，$\bar{r}_{j,i}$ 表示第 j 个物品类 g_j 所对应的社区 c_j 中用户对第 i 个物品 I_i 的评分均值，$r_{\max}$ 表示评分矩阵中的最大评分值。

通过式(2)计算出第 i 个物品 I_i 对每个物品类的隶属度，表示为 $W_i=(w_{i,1},w_{i,2},\cdots,w_{i,k})$，$0\leqslant w_{i,j}\leqslant 1(1\leqslant j\leqslant k)$。对每个物品进行计算，就可以构成一个隶属度矩阵，表示如下：

$$W=\begin{pmatrix}W_1\\W_2\\\vdots\\W_n\end{pmatrix}=\begin{pmatrix}w_{1,1},w_{1,2},\cdots,w_{1,k}\\w_{2,1},w_{2,2},\cdots,w_{2,k}\\\vdots\\w_{n,1},w_{n,2},\cdots,w_{n,k}\end{pmatrix} \tag{3}$$

3.3 物品相似度的计算

构建好物品的隶属度矩阵后，协同过滤推荐算法中的相似度将基于变换后的低维隶属度矩阵进行计算，大大缩短了传统协同过滤算法中相似度的计算时间。同时，变换后的矩阵为稠密矩阵，在一定程度上缓解了评分矩阵的稀疏性问题，因此也能够在一定程度上提高最近邻查找的精度。常用的相似度包括以下几种。

(1)基于距离的相似度：两个物品之间的相似度根据它们之间的距离来度量。定义如下：

$$sim_1(I_i,I_j)=\exp\left(-\sqrt{\sum_{q=1}^{k}|w_{i,q}-w_{j,q}|}\right) \tag{4}$$

其中，k 表示物品类的个数，$w_{i,q}$ 和 $w_{j,q}$ 分别表示第 i 个物品 I_i 和第 j 个物品 I_j 对物品类 g_q 的隶属程度。

(2)基于余弦的相似度。定义如下：

$$sim_2(I_i,I_j)=\frac{\sum_{q=1}^{k}w_{i,q}\cdot w_{j,q}}{\sqrt{\sum_{q=1}^{k}w_{i,q}}\cdot\sqrt{\sum_{q=1}^{k}w_{j,q}}} \tag{5}$$

式中各个符号的意义与式(4)相同。

(3)皮尔逊相似度。定义如下：

$$sim_3(I_i,I_j)=\frac{\sum_{q=1}^{k}w_{i,q}\cdot w_{j,q}}{\sqrt{\sum_{q=1}^{k}w_{i,q}-\bar{w}_i}\cdot\sqrt{\sum_{q=1}^{k}w_{j,q}-\bar{w}_j}} \tag{6}$$

其中，k 表示物品类的个数；$w_{i,q}$ 和 $w_{j,q}$ 分别表示第 i 个物品 I_i 和第 j 个物品 I_j 对物品类 g_q 的隶属程度；$\bar{w}_i$ 和 $\bar{w}_j$ 分别表示第 i 个物品 I_i 和第 j 个物品 I_j 对各个物品类的隶属度均值。

3.4 算法描述及时间复杂度的分析

基于上述思想及相关定义，给出基于空间变换的协同过滤推荐算法的流程，如图 2 所示。算法的具体描述如算法 1 所示。

算法 1 基于空间变换的协同过滤推荐算法

输入：评分矩阵，社交关系矩阵

输出：为目标用户产生的推荐列表

Step 1 基于社交网络信息，利用社区发现算法将用户分为不同的用户组；

Step 2 基于评分信息，根据用户和物品之间的对应关系找到各个用户类对应的物品类；

Step 3 根据式(3)构建物品隶属度矩阵；

Step 4 根据变换后的隶属度矩阵，计算物品的最近邻；

Step 5 基于目标用户对最近邻物品的评分信息，计算目标用户对目标物品的预测评分；

Step 6 选择预测评分最高的前 L 个物品推荐给目标用户。

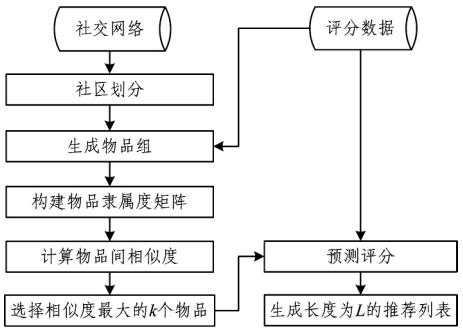

图 2 算法流程图

Fig. 2 Flowchart of proposed algorithm

本文所提算法与传统的协同过滤推荐算法的计算复杂性

之间的区别主要体现在物品相似度的计算方面。传统的协同过滤推荐算法计算物品相似度的时间复杂度为 $O(mn^2)$，其中 n 表示物品数量，m 表示用户数量。本文所提算法基于变换的隶属度矩阵进行相似度计算，时间复杂度为 $O(kn^2)$，其中 k 表示物品类个数。由于 $k \ll n$，因此时间复杂度为 $O(n^2)$。根据以上分析，与传统协同过滤推荐算法相比，本文所提算法的时间复杂度更低，具有良好的扩展性。

4　实验结果与分析

为验证本文所提算法的有效性，在公开的 Epinions 数据集上将其与已有的协同过滤推荐算法进行了比较分析。Epinions 数据集中包含 40163 位用户给 139738 个汽车、图书、电影等物品的评分信息，通常用数值 1～5 表示，该评分数据的稀疏度为 99.99%。此外，该数据集还包含了 442979 条用户之间的信任关系，如果信任，标记为 1，否则标记为 0，信任数据的稀疏度为 99.97%。实验环境为：4 GB 内存，Intel (R) Core(TM) i7-2600 处理器，3.4 GHz，Windows 7 操作系统。

4.1　评价指标

本文采用了两类指标对推荐结果的有效性进行评价。

第一类指标度量评分预测的误差率，即比较真实评分和预测评分之间的差距。采用平均绝对误差(MAE)和均方根误差(RMSE)[17]来评价预测的误差率。MAE 和 RMSE 的指标值越小，表示推荐效果越好。两个指标分别定义如下：

$$MAE=\frac{\sum_{i,j}|R_{ij}-R'_{ij}|}{N} \tag{7}$$

$$RMSE=\sqrt{\frac{\sum_{i,j}(R_{ij}-R'_{ij})^2}{N}} \tag{8}$$

其中，R_{ij} 表示第 i 个用户 U_i 对第 j 个物品 I_j 的实际评分，R'_{ij} 表示第 i 个用户 U_i 对第 j 个物品 I_j 的预测评分，N 表示测试集中包含的评分数量。

第二类指标度量推荐列表的分类准确性。假设测试集中评分大于 3 表示对应用户喜欢物品，推荐列表前端的预测评分值越高，则表示该推荐列表越好。本文使用平均准确率(MAP)进行度量，其值越大，表示生成的推荐列表的质量越好。具体定义如下：

$$MAP=\frac{1}{|U|}\sum_{j=1}^{|U|}\frac{1}{m_j}\sum_{k=1}^{m_j}Precision(R_{jk}) \tag{9}$$

其中，$|U|$ 表示总共推荐的用户数量，m_j 表示用户 U_j 在测试集中喜欢的物品数量，$Precision(R_{jk})$ 表示物品 I_k 在推荐列表中位置的倒数。

4.2　对比算法

为了验证本文所提算法的有效性，分别将其与以下几种算法进行比较与分析。

(1)传统的基于用户的协同过滤算法(UCF)[18]：只利用评分信息，度量用户之间的皮尔逊相似度，进而基于用户最近邻进行预测推荐。

(2)传统的基于物品的协同过滤算法(ICF)[19]：只利用评分信息，度量物品间的皮尔逊相似度，进而基于物品最近邻进行预测推荐。

(3)基于社交网络缺失值填充的协同过滤推荐算法(SNCF)[15]：在基于物品的协同过滤推荐框架下，分别在物品相似度的计算阶段和用户对物品的评分预测阶段利用社交网络中的朋友关系信息对评分矩阵中的缺失值进行选择性填充，然后进行推荐。

(4)融合朋友和相似用户的协同过滤算法(CNCF)[13]：利用评分信息和社交关系信息，由社交关系强度最大的用户和评分相似度最大的用户共同构成传统最近邻，然后采用与 UCF 相同的策略进行预测推荐。

(5)基于用户朋友的协同过滤算法(FCF)[13]：由社交关系强度最大的用户构成最近邻，然后采用与 UCF 相同的策略进行预测推荐。

4.3　不同训练集比例下隐式社交关系填充阈值对算法的影响

当社交网络中用户间的评分信息的皮尔逊相似度大于一定的阈值 μ 时，在用户间建立隐式的社交关系。图 3 展示了当阈值 μ 分别取 0.5 和 0.8 时在社交网络中加入隐式社交关系以及不增加隐式社交关系 3 种情况下，本文所提算法的推荐精度随训练集比例变化的情况。

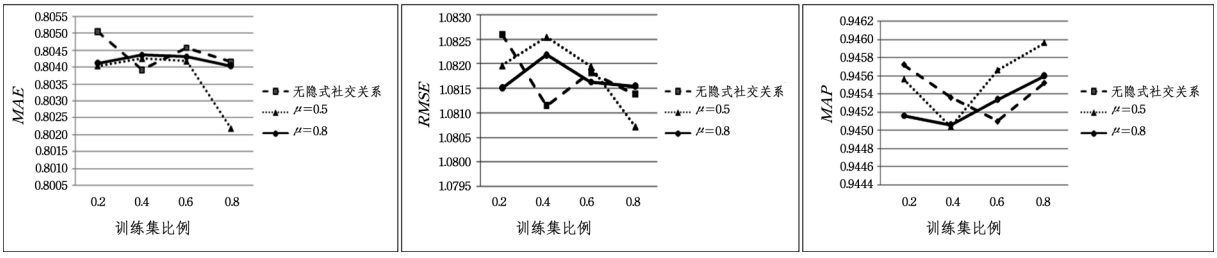

图 3　不同训练集比例对所提算法推荐精度的影响

Fig. 3　Impact of different training set ratios on recommendation accuracy of proposed algorithm

从图 3 中可以看出，当训练集比例达到 0.6 以上时，隐式社交关系的补充可以有效提高算法的预测精度。这主要是由于当训练集比例达到 0.6 以上时，用户之间的评分相似度计算有了一定的评分数量作为基础，可以更加准确地找到兴趣相似的用户对，从而提高社区发现的准确度。相反，当训练集比例太低时，由于评分信息数量太少，不足以保证用户之间相似度结果的可靠性，使得社交网络中加入了噪声，可能会干扰用户的社区发现过程，进而影响最终的推荐精度。同时，我们发现，当用户之间的评分相似度阈值设定为 0.5 时，加入隐式社交关系对算法效果的提升最明显。而当阈值设定为 0.8 时，算法的结果略有改善，这是因为相似度大于 0.8 的用户对较少，会使得加入的隐式社交关系也较少，对社区划分的改善

程度非常有限。同时,当阈值过小时,会加入很多本来就不该存在的社交关系,反而会降低社区划分的准确度。因此,在本文中,选择相似度阈值 $\mu=0.5$,如果两个用户的评分信息的相似度大于该阈值且不存在显式的社交关系,则对该用户对增加隐式社交关系。

4.4 社区数量对算法的影响

从图 4 中可以看到,当社区数量为 100 时,所提算法在 3 个指标上都达到了最优值,继续增大社区数量对结果没有更大的改善,反而会引入很多小社区。这些小社区数量很多,但包括的用户数量比大社区少得多。因此,当社区数量选择得过大时,会降低物品组划分的质量,影响算法的推荐精度。同时,当社区数量过少时,用户覆盖度很低,导致物品组划分的重叠度也很低,本来隶属于多个组的物品可能会退化为单一物品组,降低了物品的相似度计算精度。因此,本文设置用户社区个数为 100。

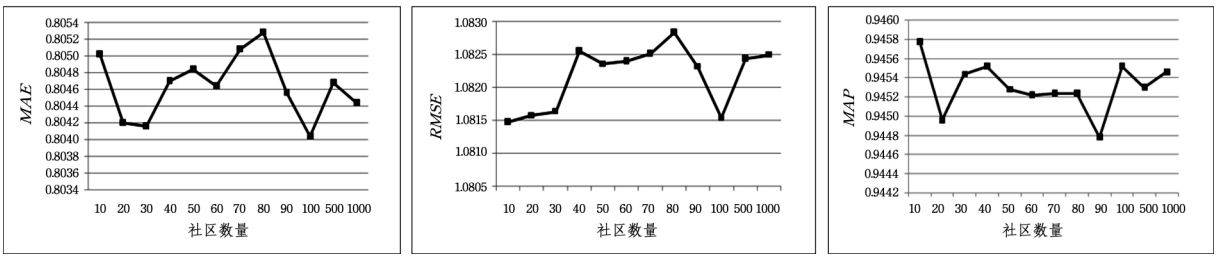

图 4 不同社区数量对所出算法推荐精度的影响

Fig. 4 Impact of different number of communities on recommendation accuracy of proposed algorithm

4.5 相似度指标对结果的影响

本节通过实验分析了文中所描述的 3 种物品相似度计算方法(基于余弦的相似度、皮尔逊相似度和基于距离的相似度)对所提算法推荐精度的影响,结果如图 5 所示。从中可以看出,基于余弦的相似度在 MAE,RMSE 和 MAP 3 种评价指标上都有明显的优势,皮尔逊次之,基于距离的相似度结果最差。基于皮尔逊和基于距离两类相似度的推荐效果差距较小。因此,本文采用余弦相似度来计算物品间的相似程度。

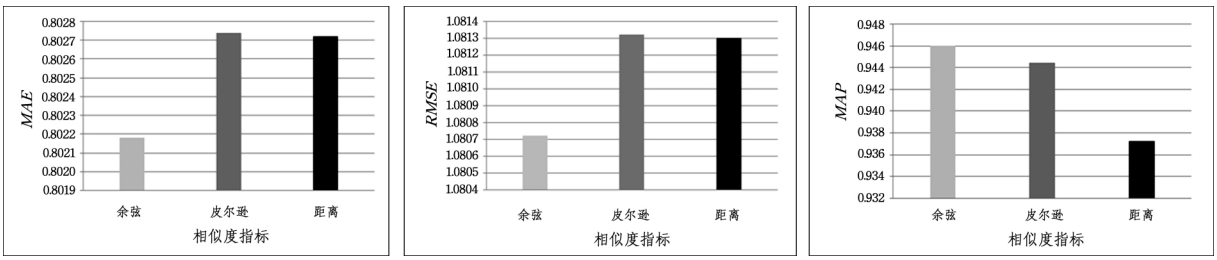

图 5 不同相似度计算方法对所提算法推荐精度的影响

Fig. 5 Impact of different similarity computation methods on recommendation accuracy of proposed algorithm

4.6 与其他推荐算法的比较

本文算法与其他推荐算法在 3 个评价指标上的实验结果如图 6 所示。需要说明的是,在本文算法中,选择的社区数量为 100,利用皮尔逊相似度度量用户的评分相似性,隐式社交关系的相似度阈值为 0.5,基于余弦的相似度度量物品隶属度向量的相似度,物品的邻居数量为 30。本文所有实验均采用五折交叉验证方法。由图 6 可知,本文所提算法在 3 个指标上均取得了最优结果。

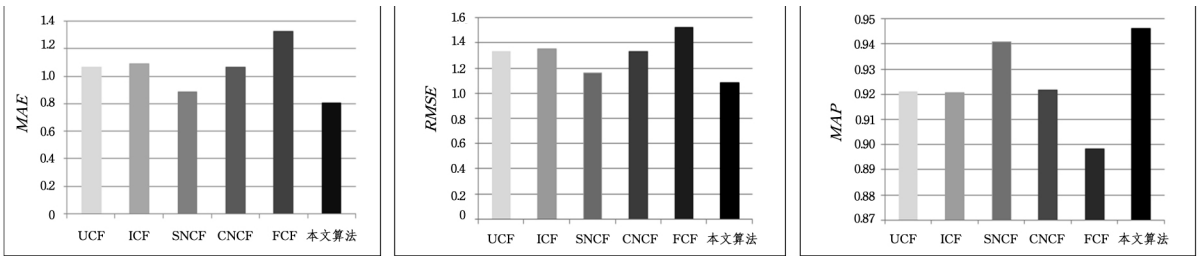

图 6 不同算法推荐精度的比较

Fig. 6 Comparison of recommendation accuracy of different algorithms

4.7 算法运行时间的比较

为了验证基于物品的协同过滤推荐算法在可扩展性上的提升,对比了所提算法与传统的协同过滤算法在不同训练集/测试集比例下的运行时间,结果如图 7 所示。运行时间包括物品间的相似度计算、相似度大小排序和预测评分 3 个阶段的运行时间。由图 7 可以看出,随着测试集比例的增大,两种

算法的运行时间都在增加，且传统的协同过滤算法耗时更长，时间增长幅度更大。这是因为随着测试集比例的增加，算法需要预测的评分数量也在增加，因此耗时会增加。但本文所提算法的时间消耗和增幅要远小于传统的协同过滤推荐算法，这是因为所提算法是基于低维的物品隶属度矩阵来计算的。

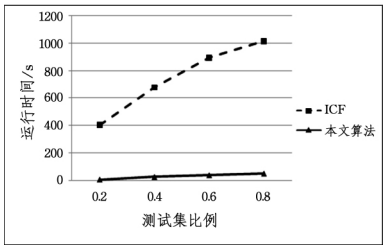

图 7　所提算法与传统的协同过滤算法在不同测试集比例下的预测评分时间对比

Fig. 7　Comparison of running time for predicting scores between proposed algorithm and traditional collaborative filtering algorithm under different test set ratios

结束语　本文提出了一种基于空间变换的协同过滤推荐算法，首先根据社交网络的显式关系和隐式关系对用户进行划分，然后通过映射得到相应的物品类，通过变换得到物品对物品类的隶属度矩阵，并通过该矩阵来计算物品间的相似度。由于物品的表示从高维稀疏的评分向量转变成了低维稠密的隶属度向量，降低了算法的时间复杂度，使得算法的可扩展性得到增强。最后，在公开数据集上进行了实验分析，验证了所提算法的有效性。

参 考 文 献

[1] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.

[2] ZHU Y Y, SUN J. Recommender system: Up to now [J]. Journal of Frontiers of Computer Science and Technology, 2015, 9(5): 513-525. (in Chinese)
朱扬勇，孙婧. 推荐系统研究进展[J]. 计算机科学与探索，2015，9(5)：513-525.

[3] LENG Y J, LU Q, LIANG C Y. Survey of Recommendation Based on Collaborative Filtering [J]. Pattern Recognition and Artificial Intelligence, 2014, 27(8): 720-734. (in Chinese)
冷亚军，陆青，梁昌勇. 协同过滤推荐技术综述[J]. 模式识别与人工智能，2014，27(8)：720-734.

[4] GOLDBERG D, NICHOLS D, OKI B M, et al. Using collaborative filtering to weave an information tapestry [J]. Communications of the ACM, 1992, 35(12): 61-70.

[5] SHI Y, LARSON M, HANJALIC A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges[J]. ACM Computing Surveys, 2014, 47(1): 1-45.

[6] LENG Y J, LIANG C Y, DING Y, et al. Method of neighborhood formation in collaborative filtering[J]. Pattern Recognition and Artificial Intelligence, 2013, 26(10): 968-974. (in Chinese)
冷亚军，梁昌勇，丁勇，等. 协同过滤中一种有效的最近邻选择方法[J]. 模式识别与人工智能，2013，26(10)：968-974.

[7] WANG J, VRIES A P D, REINDERS M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion[C]// 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2006.

[8] LIANG C, LENG Y. Collaborative filtering based on information-theoretic co-clustering[J]. International Journal of Systems Science, 2014, 45(3): 589-597.

[9] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. IEEE Computer, 2009, 42(8): 30-37.

[10] ZENG W, ZENG A, LIU H, et al. Uncovering the information core in recommender systems[J]. Scientific Reports, 2014, 4: 6140.

[11] CAI Y, LEUNG H, LI Q, et al. Typicality-based collaborative filtering recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 2(3): 97-104.

[12] XU B, BU J, CHEN C, et al. An exploration of improving collaborative recommender systems via user-item subgroups[C]// International Conference on World Wide Web. 2012: 21-30.

[13] LIU F, HONG J H. Use of social network information to enhance collaborative filtering performance[J]. Expert Systems with Applications, 2010, 37(7): 4772-4778.

[14] GUO G B, ZHANG J, THALMANN D. Merging trust in collaborative filtering to alleviate data sparsity and cold start[J]. Knowledge-Based Systems, 2014, 57(2): 57-68.

[15] GUO L J, LIANG J Y, ZHAO X W. Collaborative filtering recommendation algorithm incorporating social network information[J]. Pattern Recognition and Artificial Intelligence, 2016, 29(3): 281-288. (in Chinese)
郭兰杰，梁吉业，赵兴旺. 融合社交网络信息的协同过滤推荐算法[J]. 模式识别与人工智能，2016，29(3)：281-288.

[16] WALTMAN L, ECK N J V. A smart local moving algorithm for large-scale modularity-based community detection[J]. European Physical Journal B, 2013, 86(11): 1-14.

[17] TANG J, HU X, LIU H. Social recommendation: A review[J]. Social Network Analysis & Mining, 2013, 3(4): 1113-1133.

[18] BREESE J S, HECKERMAN D, KADIE C. Empirical analysis of predictive algorithms for collaborative filtering[C]// 14th Conference on Uncertainty in Artificial Intelligence. 1998: 43-52.

[19] DESHPANDE M, KARYPIS G. Item-based top-n recommendation algorithms[J]. ACM Transactions on Information Systems, 2014, 22(1): 143-177.