

# 基于信任关系下用户兴趣偏好的协同过滤推荐算法

邵超 宋淑米

河南财经政法大学计算机与信息工程学院 郑州 450046

**摘要** 随着信息的海量增长,推荐系统有效缓解了信息爆炸带来的问题,其中协同过滤作为主流技术之一受到了广泛的关注。针对用户的兴趣偏好研究主要是基于商品标签的有监督数据集进行研究,忽略了无监督数据集,同时,在计算用户的兴趣偏好过程中也未能考虑到信任用户对用户兴趣的影响。为此,文中首先在无监督的项目数据集上采用矩阵分解模型得到项目的潜在特征向量,据此对项目进行聚类以表示项目的类别信息;然后,结合用户的信任关系和用户-项目评分矩阵构造用户的兴趣偏好矩阵;最后,为提高推荐效率,在用户的兴趣偏好矩阵上对用户进行聚类,在每个聚类簇内计算用户之间的相似度,从而实现推荐。在公开数据集上的实验结果表明,该算法能有效改善推荐结果的精确性,提升推荐质量。

**关键词:** 协同过滤推荐;聚类;偏好矩阵;信任关系;矩阵分解;相似度

**中图法分类号** TP181

## Collaborative Filtering Recommendation Algorithm Based on User Preference Under Trust Relationship

SHAO Chao and SONG Shu-mi

School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China

**Abstract** With the massive increase of information, the recommendation system has effectively alleviated the problems caused by the information explosion. Collaborative filtering, as one of the mainstream technologies of recommendation system, has been widely concerned. In the research of users' interest preference, the supervised data sets based on commodity labels are mainly studied, and the unsupervised data sets are ignored. At the same time, the influence of trusted users on users' interest is not considered in the process of calculating users' interest preference. To solve these problems, a collaborative filtering recommendation algorithm based on user preference under trust relationship is proposed in this paper. Firstly, the potential feature information of the items is obtained using the matrix factorization (MF) model, and then is clustered to obtain item type information. Secondly, the user trust relationship and users-item rating information are considered to construct the user preference matrix. Finally, the users are clustered based on the user preference matrix, and then the similarities between users in one cluster are calculated to implement recommendation. Experimental results on open datasets show that the algorithm can effectively improve the accuracy of recommendation results and the quality of recommendations.

**Keywords** Collaborative filtering recommendation, Clustering, Preference matrix, Trust relationship, Matrix factorization, Similarity

### 1 引言

随着信息的爆炸式增长,依赖手工输入查询条件的搜索引擎已经不能满足人们的多样化需求。与此同时,不需要用户输入精确条件便可为用户筛选出感兴趣内容的推荐系统得到了越来越多的关注,成为解决信息过载问题的重要手段。目前,推荐系统已广泛应用于淘宝、京东等电商平台和头条、抖音等 App。主要推荐技术有 Demographic filtering<sup>[1]</sup>、基于内容的推荐(Content Based)<sup>[2]</sup>、协同过滤推荐(Collaborate Filtering, CF)<sup>[3]</sup>等。其中,应用最广泛的推荐技术是协同过滤推荐,它不需要用户的基本信息和项目自身的信息,只依据用户的历史行为(浏览、评分、评价等)进行推荐。

协同过滤推荐又可划分为基于内存(Memory-based)的协同过滤推荐、基于模型(Model-based)的协同过滤推荐、混合(Hybrid-CF)协同过滤推荐。基于内存的协同过滤推荐主要有基于用户(User-based)的协同过滤推荐<sup>[4]</sup>和基于项目(Item-based)的协同过滤推荐<sup>[5]</sup>。该方法简单易实现,通过计算用户或项目之间的相似度进行推荐,但在整个数据集上计算相似度存在数据稀疏和维度过高等问题。基于模型的协同过滤推荐通过创建一个模型解决了基于内存的协同过滤推荐存在的上述问题,主要有基于贝叶斯模型(Baysian Models)<sup>[6]</sup>的协同过滤推荐、基于聚类(Clustering)<sup>[7]</sup>的协同过滤推荐、基于矩阵分解(Matrix Factorization, MF)<sup>[8]</sup>的协同过滤推荐、基于回归(Regression)<sup>[9]</sup>的协同过滤推荐、基于 Markov 模

基金项目:国家自然科学基金(61202285,61502146,61841702)

This work was supported by the National Natural Science Foundation of China(61202285,61502146,61841702).

通信作者:邵超(sc\_flying@163.com)

型<sup>[10]</sup>的协同过滤推荐等方法。与基于内存的协同过滤推荐相比,这些模型取得了较好的推荐效果。混合协同过滤推荐主要将以上两种协同过滤推荐进行了融合。

然而,只依赖用户-项目评分矩阵的推荐系统容易受到稀疏矩阵的干扰。为降低数据的稀疏性,通过挖掘用户的兴趣偏好,人们又提出了基于用户兴趣偏好的协同过滤推荐。Huang等<sup>[11]</sup>提出了结合用户兴趣度聚类的协同过滤推荐算法,通过提取项目属性中的关键词,构造用户-关键词偏好矩阵,并在偏好矩阵上进行聚类。Zhu等<sup>[12]</sup>提出了基于标签聚类和用户兴趣划分的协同过滤推荐算法,通过对项目标签进行聚类,完成用户兴趣度的划分,然后在兴趣度上进行推荐。He等<sup>[13]</sup>提出了聚类和用户兴趣偏好相融合的协同过滤推荐算法,通过项目的标签构造用户的兴趣偏好矩阵,然后在用户的兴趣偏好矩阵上对用户进行聚类。这些研究不论是根据关键词还是根据项目标签来构造用户偏好矩阵,都是在有监督的项目数据集上进行的,未能对无监督的数据集进行研究。Wang等<sup>[14]</sup>通过用户之间的相似性来构造用户对具体项目的偏好,然后结合用户信任机制得到推荐系数,以提高推荐的准确性。Wang等<sup>[15]</sup>提出一种结合用户之间加权信任网络与相似性的社会化模型,将用户的信任关系进行细分,提高了推荐系统的性能。Wang等<sup>[16]</sup>根据信任的传递机制,提出了一种TPCF算法,该算法是通过用户的评分数据计算用户间的偏好相似性,再结合信任传递机制获取任意用户之间的信任度,然后动态结合两部分指标得到综合推荐权重来提高推荐的效果。以上考虑用户信任关系的研究都是依靠用户之间的相似度或用户对具体项目的评分来判断用户的兴趣,未考虑到项目的类型信息。这也说明用户的偏好不仅和用户的评分信息相关,而且在一定程度上和用户的社交信息如用户的信任关系也相关。因此,本文首先采用矩阵分解模型在无监督的项目数据集上获取项目的潜在特征向量,据此对项目进行聚类以获得项目的类型;然后根据项目的类型和用户-项目评分矩阵以及用户的信任关系构造用户的兴趣偏好矩阵;最后在用户的兴趣偏好矩阵上对用户进行聚类,在目标用户所属的聚类簇内计算用户之间的相似度,实现Top-N推荐。

本文的主要贡献有两个方面:

- 1)在无监督的项目数据集上计算用户的兴趣偏好。
- 2)将用户的信任关系融入到用户的兴趣偏好中。

## 2 相关知识

### 2.1 相似度

在对用户进行推荐时所依据的一个重要指标就是用户或项目之间的相似度。推荐系统主要使用3种相似度:皮尔逊相关系数(PCC)、余弦相似性、修正的余弦相似性。本文参考文献<sup>[17]</sup>,采用皮尔逊相关系数,如式(1)所示:

$$\text{sim}(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

其中, $r_{u,i}$ 表示用户 $u$ 对项目 $i$ 的评分, $\bar{r}_u$ 表示用户 $u$ 评价过的项目的平均分, $I_{uv}$ 表示用户 $u$ 和用户 $v$ 共同评价过的项目。

### 2.2 预测评分

在对用户进行个性化推荐时,需要预测出目标用户未评

分项的评分,根据预测出的评分为用户推荐评分最高的项目。本文依据的预测评分公式参考文献<sup>[13]</sup>,具体如式(2)所示。

$$\hat{r}_{u,i} = \bar{r}_u + \frac{1}{\sum_{v \in N_u} \text{sim}(u, v)} \sum_{v \in N_u} \text{sim}(u, v)(r_{v,i} - \bar{r}_v) \quad (2)$$

其中, $\hat{r}_{u,i}$ 表示用户 $u$ 对项目 $i$ 的预测评分, $N_u$ 表示与用户 $u$ 最相似的 $k$ 个用户的集合。

### 2.3 矩阵分解

矩阵分解是将一个高维稀疏性矩阵转换为两个低维矩阵乘积的形式,从而找到用户和项目的潜在特征向量。本文参考文献<sup>[8]</sup>的处理方法,最小化如式(3)所示的目标函数。

$$L = \min_{p, q} \frac{1}{2} \sum_{(u, i) \in R} \|r_{u,i} - p_u q_i^T\|^2 + \frac{1}{2} \lambda (\|p_u\|^2 + \|q_i\|^2) \quad (3)$$

其中, $p_u$ 表示用户 $u$ 的潜在特征向量, $q_i$ 表示商品 $i$ 的潜在特征向量, $\lambda$ 为正则化参数。本文采用随机梯度下降算法最小化目标函数 $L$ ,从而得到最优的 $p_u$ 和 $q_i$ 。

## 3 信任关系下用户兴趣偏好的协同过滤推荐算法

用户的兴趣和项目的类型密切相关,但一些数据库并没有明确给出项目的类型,本文采用矩阵分解模型求出项目的潜在特征向量,并据此对项目进行聚类,不同的聚类簇代表不同的项目类型。然后结合用户-项目评分矩阵和用户的信任关系构造用户的兴趣偏好矩阵。最后对用户的兴趣偏好矩阵进行聚类,求出目标用户所属聚类簇中用户之间的相似度,找到与目标用户最相似的 $k$ 个用户,根据预测评分公式进行评分预测,为目标用户推荐评分最高的项目。本文算法的整体框架如图1所示。

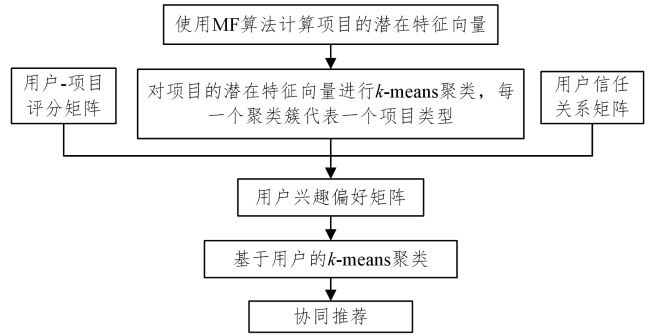


图1 本文算法框架

Fig. 1 Algorithm framework of this paper

### 3.1 基于信任关系的用户兴趣偏好矩阵

协同过滤推荐算法仅仅考虑了用户的历史评分信息,忽略了项目的特征,但是用户的兴趣和项目的特征存在着紧密的联系。因此,He等<sup>[13]</sup>利用商品的标签信息,构造了用户-项目类型(User-Type, UT)的用户偏好矩阵,采用如式下公式:

$$P(u)_x = \frac{\sum_{i \in I_{u,x}} r_{u,i}}{\text{sum}(I_{u,x})} \quad (4)$$

其中, $P(u)_x$ 表示用户 $u$ 对项目标签为 $x$ 的项目类别偏好, $I_{u,x}$ 表示用户 $u$ 评价过的项目中标签为 $x$ 的项目集合, $\text{sum}(I_{u,x})$ 表示用户 $u$ 评价过的项目中标签为 $x$ 的项目个数。

虽然用户的兴趣偏好和项目特征密切相关,但是项目特

征并不是唯一影响用户兴趣的因素,用户之间的信任关系也会影响用户的兴趣偏好。例如一个用户只喜欢看教育片,不喜欢看爱情片,但是他看到自己信任的朋友和影评人都推荐看这部爱情电影,那这个用户很可能会去看这部爱情片。因此,本文在构造用户的兴趣偏好矩阵时,将用户的信任关系考虑进去,构造基于用户信任关系的兴趣偏好矩阵(Trust-UT),如式(5)所示:

$$P(u)_t = \alpha \frac{\sum_{i \in I_{u,t}} r_{u,i}}{\text{sum}(I_{u,t})} + (1-\alpha) \frac{\sum_{u' \in L_u} P(u')_t}{\text{sum}(L_u)} \quad (5)$$

其中, $P(u)_t$ 表示用户 $u$ 对类型 $t$ 的项目的偏好, $I_{u,t}$ 表示用户 $u$ 评价过的项目中属于类型 $t$ 的项目集合, $L_u$ 表示用户 $u$ 的信任用户集合, $\text{sum}(\cdot)$ 表示集合中的元素个数, $\alpha$ 为平衡因子。

例如,现有用户 $u_1, u_2, u_3, u_4$ 对项目 $i_1, i_2, i_3, i_4, i_5$ 的评分矩阵 $\mathbf{R}$ 如表1所列,项目 $i_1, i_2, i_3, i_4, i_5$ 所属的聚类簇矩阵 $\mathbf{T}$ 如表2所列,用户 $u_1, u_2, u_3, u_4$ 的信任关系矩阵 $\mathbf{F}$ 如表3所列。

表1 用户-项目评分矩阵 $\mathbf{R}$   
Table 1 User-item rating matrix  $\mathbf{R}$

	$i_1$	$i_2$	$i_3$	$i_4$	$i_5$
$u_1$	2	4	2	0	0
$u_2$	0	3	2	0	5
$u_3$	4	0	5	1	0
$u_4$	2	0	4	1	3

表2 项目类型矩阵 $\mathbf{T}$   
Table 2 Item type matrix  $\mathbf{T}$

	$t_1$	$t_2$
$i_1$	1	0
$i_2$	0	1
$i_3$	0	1
$i_4$	1	0
$i_5$	0	1

表3 用户信任关系矩阵 $\mathbf{F}$   
Table 3 Users' trust relationship matrix  $\mathbf{F}$

	$u_1$	$u_2$	$u_3$	$u_4$
$u_1$	0	1	1	0
$u_2$	1	0	0	1
$u_3$	1	1	0	1
$u_4$	0	0	1	0

表1中, $r_{u,i}=0$ 表示用户 $i$ 未对项目 $j$ 进行评分。表2中, $t_{i,j}=1$ 表示项目 $i$ 属于聚类簇 $j$ , $t_{i,j}=0$ 表示项目 $i$ 不属于聚类簇 $j$ 。表3中, $f_{u,u'}=1$ 表示用户 $u$ 信任用户 $u'$ , $f_{u,u'}=0$ 表示用户 $u$ 不信任用户 $u'$ 。

根据以上3张表,构造信任关系下用户偏好矩阵的具体步骤如下。

第1步 计算未添加信任关系下用户对各个项目类型的偏好。

根据表1,用户 $u_1$ 评价的所有项目评分为 $i_1=2, i_2=4, i_3=2$ ;根据表2,项目 $i_1$ 属于类型 $t_1$ ,项目 $i_2$ 和 $i_3$ 属于类型 $t_2$ ,所以用户 $u_1$ 对类型 $t_1$ 的总评分为2,评分项目个数为1,对类型 $t_2$ 的总评分为6,评分项目个数为2,根据式(4),用户 $u_1$ 对类型 $t_1$ 的偏好为2,对类型 $t_2$ 的偏好为3。同理,用户 $u_2, u_3, u_4$ 的偏好如表4所列。

表4 用户的偏好矩阵 $\mathbf{UT}$

Table 4 Users' preference matrix  $\mathbf{UT}$

	$t_1$	$t_2$
$u_1$	2	3
$u_2$	0	3.33
$u_3$	2.5	5
$u_4$	1.5	3.5

第2步 计算信任用户的偏好。

由表3可知,用户 $u_1$ 的信任用户是 $u_2, u_3$ ,他们对类型 $t_1$ 的偏好总评分为2.5,平均评分为1.25;对类型 $t_2$ 的偏好总评分为8.33,平均评分为4.17。同理,用户 $u_2$ 的信任用户对类型 $t_1$ 的偏好平均评分为1.75,对类型 $t_2$ 的偏好平均评分为3.25;用户 $u_3$ 的信任用户对类型 $t_1$ 的偏好平均评分为1.17,对类型 $t_2$ 的偏好平均评分为3.28;用户 $u_4$ 的信任用户对类型 $t_1$ 的偏好平均评分为2.5,对类型 $t_2$ 的偏好平均评分为5。

第3步 计算信任关系下用户兴趣偏好矩阵(Trust-UT)。

根据式(5),假设 $\alpha=0.7$ ,则用户 $u_1$ 在信任关系下对类型 $t_1$ 的兴趣偏好为 $0.7 \times 2 + (1-0.7) \times 1.25 = 1.78$ ,对类型 $t_2$ 的兴趣偏好为 $0.7 \times 3 + (1-0.7) \times 4.17 = 3.35$ 。同理,用户 $u_2, u_3, u_4$ 在信任关系下的兴趣偏好如表5所列。

表5 信任关系下用户兴趣偏好矩阵(Trust-UT)

Table 5 User interest preference matrix under trust relationships

Trust-UT		
	$t_1$	$t_2$
$u_1$	1.78	3.35
$u_2$	0.53	3.31
$u_3$	2.1	4.48
$u_4$	1.8	3.95

综上所述,信任关系下用户兴趣偏好矩阵的构造过程如算法1所示。

### 算法1 信任关系下用户兴趣偏好提取算法

输入: 用户-项目评分矩阵 $\mathbf{R}$ ,项目类型矩阵 $\mathbf{T}$ ,信任关系矩阵 $\mathbf{F}$ ,平衡因子 $\alpha$

输出: 信任关系下用户兴趣偏好矩阵 Trust-UT

Step 1 利用矩阵乘法计算用户对每个项目类型的总评分;

Step 2 计算每个项目类型下用户评分的项目个数;

Step 3 根据式(4)计算未考虑信任关系的用户兴趣偏好矩阵;

Step 4 计算用户的信任用户个数及其对每个项目类型的偏好总评分;

Step 5 根据式(5)计算信任关系下用户的兴趣偏好矩阵。

### 3.2 聚类

在传统的协同过滤推荐算法中,计算用户或项目之间的相似度是在整个数据集上进行的,这种方法的效率比较低,还存在数据稀疏等问题。为解决上述问题,一些学者提出采用聚类和协同过滤相结合的协同聚类算法。协同聚类算法考虑到用户或项目之间存在的某种相关性,首先将用户或项目进行聚类,然后在每个聚类簇内计算用户或项目之间的相似度。

本文使用了两次聚类算法:第一次在用MF算法求出项目的潜在特征向量后,根据潜在特征向量对项目进行 $k$ -means聚类(聚类簇个数用 $c_1$ 表示),聚类结果中的每个聚类簇可以表示一个项目类型;第二次是计算出用户的兴趣偏好矩阵 Trust-UT后,根据兴趣偏好矩阵对用户进行 $k$ -means聚类(聚类簇个数用 $c$ 表示)。

$k$ -means聚类需要预先指定聚类簇的个数,即 $c_1$ 和 $c$ ,这将在实验部分进行详细说明。

### 3.3 算法流程

综上所述,本文提出的基于信任关系下用户兴趣偏好的协同过滤算法(Trust-UTCF)的具体流程如算法2所示。

**算法2** 基于信任关系下用户兴趣偏好的协同过滤算法(Trust-UTCF)

输入:用户评分记录(rating.txt),用户之间信任关系记录(trust.txt),  
参数  $c_1, c, \alpha, k$

输出:预测评分矩阵  $\hat{R}_{n \times m}$

Step 1 数据预处理。分别根据用户评分记录(rating.txt)和用户之间信任关系记录(trust.txt)构造用户评分矩阵  $R_{n \times m}$  和用户信任关系矩阵  $F_{n \times n}$ ,  $n$  为用户数,  $m$  为项目数。

Step 2 求解项目的潜在特征向量。采用矩阵分解模型求出项目的潜在特征向量,其中,特征向量维度  $D=10$ ,学习率 learning\_rate=0.001,正则化参数  $\lambda=0.1$ ,最大迭代次数 max\_iteration=100。

Step 3 对项目的潜在特征向量进行 k-means 聚类,其中,聚类簇个数设置为  $c_1$ 。

Step 4 构造项目类型矩阵  $T_{m \times c_1}$ 。

Step 5 根据式(5)构造信任关系下用户兴趣偏好矩阵 Trust-UT  $n \times c_1$ ,其中平衡因子设置为  $\alpha$ 。

Step 6 对 Trust-UT  $n \times c_1$  矩阵进行 k-means 聚类,其中,聚类簇个数设置为  $c$ 。

Step 7 根据式(2)选取近邻用户数  $k$ ,在每个聚类簇内计算用户对项目的预测评分,最后形成用户对项目的预测评分矩阵  $\hat{R}_{n \times m}$ 。

Step 8 将预测评分最高的项目推荐给用户。

## 4 实验结果与分析

### 4.1 实验设置

为验证本文算法(Trust-UTCF)的有效性,本文将在两个公开数据集 FilmTrust 和 Epinions 上对 Trust-UTCF 和 5 个基准的协同过滤算法进行对比实验。

FilmTrust 是一个小规模的电影数据集,包含了 1508 个用户对 2071 个项目的评分记录以及 1853 条用户之间的信任关系记录。Epinions 是一个中等规模的数据集,包含了 40163 个用户对 139738 个项目的 664624 条评分记录,该数据集中许多用户的评分次数过少或一些项目被用户评分的次数过少。本文对数据进行了预处理,即去掉了用户评分次数少于 30 次的用户信息以及项目被评分次数少于 30 次的项目信息,处理后的数据集信息如表 6 所列。

表 6 实验数据集

Table 6 Datasets in experiments

数据集	用户数	项目数	评分记录	信任关系	稀疏性/%
FilmTrust	1508	2071	35497	1853	98.86
Epinions	5383	3423	157887	72435	99.14

5 个基准的协同过滤算法如下。

1) User-based<sup>[4]</sup>(基于用户的协同过滤):该算法通过计算所有用户之间的相似度,将与目标用户相似的用户所喜欢的项目推荐给目标用户。

2) User-clustering<sup>[7]</sup>(基于用户聚类的协同推荐):该算法根据历史信息对用户进行聚类,然后计算每个聚类簇内用户之间的相似度,完成推荐。

3) UT<sup>[13]</sup>(基于用户兴趣偏好的协同推荐):该算法根据用户-项目评分矩阵和项目标签信息,构造用户兴趣偏好矩

阵,然后对用户进行聚类推荐。

4) CFIR<sup>[14]</sup>(结合信任机制和用户偏好的协同推荐):该算法通过建立信任机制求得每位用户的信任度,结合用户偏好获得综合推荐指标进行推荐。

5) STUI<sup>[18]</sup>(基于用户信任和兴趣的协同过滤推荐):该算法将用户信任关系信息与用户兴趣相似信息相结合,对传统协同过滤算法进行改进,提高推荐的准确性。

实验随机选取 80% 数据作为训练集,剩余 20% 数据作为测试集。需要设置的参数有 4 个:项目聚类簇个数  $c_1$ ,平衡因子  $\alpha$ ,用户聚类簇个数  $c$ ,以及近邻用户数  $k$ 。

### 4.2 实验指标

实验采用推荐系统常用的两个评价指标:平均绝对误差(MAE)和均方根误差(RMSE)对算法的推荐质量进行评估。

$$MAE = \frac{1}{|R_{test}|} \sum_{(u,i) \in R_{test}} |r_{u,i} - \hat{r}_{u,i}| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{|R_{test}|} \sum_{(u,i) \in R_{test}} (r_{u,i} - \hat{r}_{u,i})^2} \quad (7)$$

### 4.3 实验结果与参数调整

表 7 列出了 5 个基准协同过滤推荐算法和本文算法(Trust-UTCF)在两个数据集上的实验结果(10 次实验得到的平均值)。各算法参数均设置为通过交叉验证方法确定的使推荐结果最好的参数值:对于 FilmTrust 数据集,  $c_1=12$ ,  $\alpha=0.6$ ,  $c=14$ ,  $k=35$ ;对于 Epinions 数据集,  $c_1=22$ ,  $\alpha=0.8$ ,  $c=22$ ,  $k=25$ 。表 7 中,improve 代表本文算法结果较其他算法的结果所提升的比例,其计算公式是(基准算法结果-本文算法结果)/基准算法结果。

表 7 不同算法的 MAE 和 RMSE 值对比

Table 7 Comparison of MAE and RMSE of different algorithms

Algorithm	FilmTrust		Epinions	
	MAE	RMSE	MAE	RMSE
	improve	improve	improve	improve
User-based	0.6711	0.8954	1.0034	1.2872
	10.19%	11.63%	9.77%	7.47%
User-clustering	0.6348	0.8010	0.9701	1.2764
	5.06%	1.21%	6.67%	6.68%
UT	0.6217	0.8258	0.9504	1.2577
	3.06%	4.18%	4.73%	5.30%
CFIR	0.6087	0.7942	0.9069	1.1961
	0.99%	0.37%	0.17%	0.42%
STUI	0.6046	0.8012	0.9062	1.1957
	0.31%	1.24%	0.09%	0.38%
Trust-UTCF	<b>0.6027</b>	<b>0.7913</b>	<b>0.9054</b>	<b>1.1911</b>

从实验结果可以看到:与已有的 User-based, User-clustering, UT, CFIR 和 STUI 算法相比,本文提出的 Trust-UTCF 算法在两个数据集上的推荐精确度都得到了不同程度的提升,尤其是相比前 3 个算法,推荐准确性明显提高;对于 CFIR 和 STUI 算法,虽本文提出的算法也有所提升,但是提升度较低,说明信任度确实可以提升推荐的准确性,但用户信任度的度量标准也会影响推荐的结果。此外,我们还可以发现,与 FilmTrust 相比, Epinions 上的误差更高一些,这是由于 Epinions 的稀疏性更高,从而导致算法难以获取用户准确的兴趣偏好,说明数据的稀疏性对实验的影响也是比较大的。

本文算法涉及到 4 个参数,其中  $c_1$  和  $c$  是 k-means 聚类时需要设置的聚类簇个数,  $\alpha$  是计算用户兴趣偏好时为平衡用户自己及其信任用户之间的比例而设置的,  $k$  是在最后进

行推荐时选取的近邻用户数。在不同数据集上,参数的影响也是不同的,下面分别在两个数据集上采用控制变量法对不同参数进行调整。如图 2 和图 3 所示,每个子图的横轴代表不同参数的变化范围,纵轴表示不同的参数值对实验结果 MAE 和 RMSE 的影响。

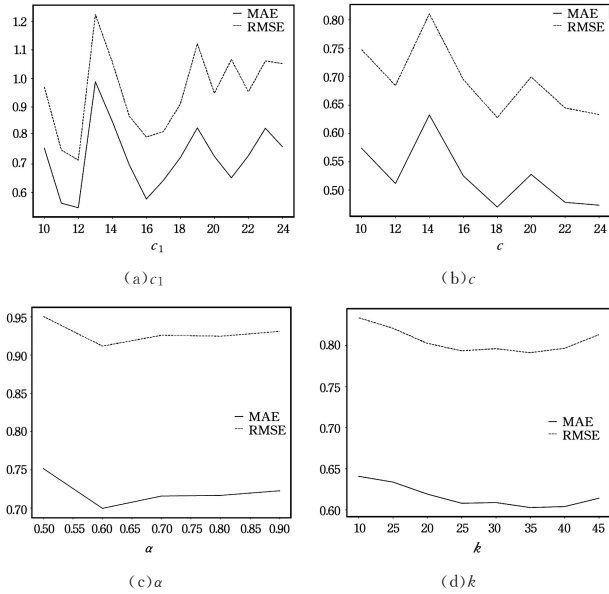


图 2 FilmTrust 数据集上参数的影响

Fig. 2 Influence of parameters on FilmTrust dataset

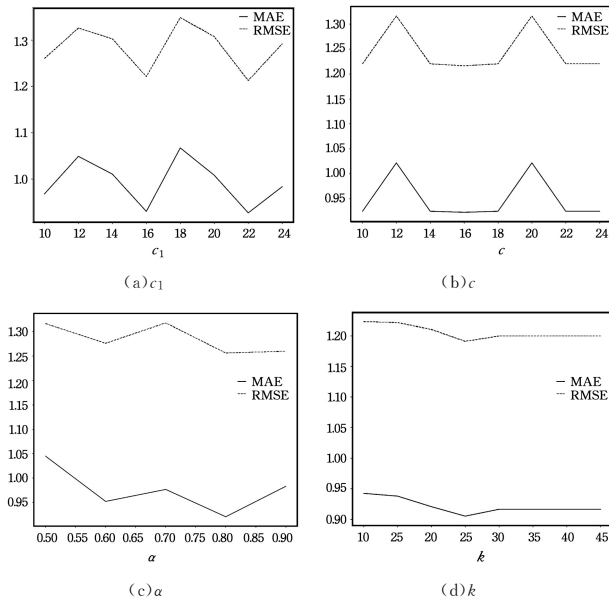


图 3 Epinions 数据集上参数的影响

Fig. 3 Influence of parameters on Epinions dataset

图 2 是 FilmTrust 数据集上 4 个参数对实验结果的影响,可以看到随着项目聚类簇个数  $c_1$  的增加,MAE 和 RMSE 都呈现大幅波动的态势,说明项目聚类簇个数对实验影响较大,在  $c_1=12$  时,MAE 和 RMSE 都处于最低的水平;用户聚类簇个数  $c$  对实验误差的影响是上升后下降,在  $c=14$  时达到偏差最大值, $c=18$  时偏差最小;平衡系数  $\alpha$  对实验结果的影响是先降低,后增高,当  $\alpha=0.6$  时,偏差最小,所以本文选取 0.6 作为最终的值,这也说明了用户的信任关系对用户的兴趣影响还是较大的;近邻用户  $k$  对实验结果的影响是预测偏差先降低随后升高, $k=35$  时 MAE 和 RMSE 都最小,所以

选择 35 作为近邻用户数。

图 3 是 Epinions 数据集上 4 个参数对实验结果的影响,整体来看,这些参数在该数据集上不如在 FilmTrust 数据集上的影响大,但还是存在一定的偏差。从图中可以看到, $c_1$  对预测评分的影响也波动较大,在  $c_1=22$  时偏差最小; $c$  值对 Epinions 数据集的影响呈现出阶段性的状态,先波动后平稳,在 14~18 之间是平稳的,误差也较小,在 22~24 的范围内也是处于最小状态,因数据集较大,本文选取  $c=22$ ;  $\alpha$  对实验结果的影响也较大,在 0.8 时误差最小; $k=25$  时,MAE 最小。

**结束语** 本文采用矩阵分解模型在无监督的项目数据集上得到项目的潜在特征向量,据此对项目进行  $k$ -means 聚类,得到用户-项目类型偏好矩阵,然后将用户的信任关系融入到用户的兴趣偏好矩阵中,从而实现基于信任关系下用户兴趣偏好的协同过滤推荐。在两个公开数据集 FilmTrust 和 Epinions 上的实验表明,本文算法可以较好地提高推荐的精确度。

本文在实验中设置了较多的参数,对实验造成了较大影响,参数的设置问题将是今后研究的一个切入点,例如,采用聚类集成<sup>[19]</sup>、模型选择等方法消除聚类簇个数对实验的影响,采用集团序<sup>[20]</sup>等方法消除近邻用户数对实验的影响。另外,本文虽然考虑了用户的信任关系,但相对简单,现实生活中用户的信任机制<sup>[21]</sup>较为复杂,信任关系也是分等级的,如完全信任、较信任、不信任等,信任程度不同,对用户的影响也是不同的,这也是以后研究的一个切入点。

## 参考文献

- [1] PAZZANI M J. A Framework for Collaborative, Content-Based and Demographic Filtering [J]. Artificial Intelligence Review, 1999, 13(5/6): 393-408.
- [2] MEYMANDPOUR R, DAVIS J G. A semantic similarity measure for linked data: An information content-based approach [J]. Knowledge-Based Systems, 2016, 109(OCT. 1): 276-293.
- [3] 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.
- [4] RESNICK P. GroupLens: An open architecture for collaborative filtering of Netnews [C] // Proc. Cscw. 1994.
- [5] SARWAR B. Item-Based Collaborative Filtering Recommendation Algorithms [C] // Proc. the 10th International World Wide Web Conference (WWW10). Hong Kong, 2001.
- [6] PAN W, CHEN L. GBPR: group preference based Bayesian personalized ranking for one-class collaborative filtering [C] // Twenty-third International Joint Conference on Artificial Intelligence. 2013: 2691-2697.
- [7] ERICSON K, PALLICKARA S. On the performance of high dimensional data clustering and classification algorithms [J]. Future Generation Computer Systems, 2013, 29(4): 1024-1034.
- [8] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems [J]. IEEE Computer Journal, 2009, 42(8): 30-37.
- [9] VUCETIC S, OBRADOVIC Z. Collaborative Filtering Using a Regression-Based Approach [J]. Knowledge and Information Systems, 2005, 7(1): 1-22.

- [10] LI X. Mean-variance optimization of discrete time discounted Markov decision processes[J]. *Automatica*, 2018, 88: 76-82.
- [11] HUANG X Y, LONG S Y, XIE J. Collaborative filtering recommendation algorithm combined with user interest clustering[J]. *Application Research of Computers*, 2019, 36(9): 2609-2612, 2617.
- [12] ZHU D J, LI J Z, TAN D Y. Collaborative filtering recommendation algorithm based on label clustering and interest division[J]. *Computer Engineering*, 2017, 43(11): 146-151, 160.
- [13] HE M, SUN W, XIAO R, et al. Collaborative filtering recommendation algorithm combining clustering and user preference[J]. *Computer Science*, 2017, 44(S2): 391-396.
- [14] WANG Q, WANG J H. Collaborative filtering recommendation algorithm combining trust mechanism with user preferences[J]. *Computer Engineering and Applications*, 2015, 51(10): 261-265, 270.
- [15] WANG M, MA J. A novel recommendation approach based on users' weighted trust relations and the rating similarities[J]. *Soft Computing*, 2015, 20(10).
- [16] WANG J L. Research on the recommendation algorithm based on trust relationship and preference[D]. Xi'an: Xidian University, 2018.
- [17] PENG W, XIN B. SPMF: A social trust and preference segmentation-based matrix factorization recommendation algorithm[J]. *Eurasip Journal on Wireless Communications & Networking*, 2019: 272-284.
- [18] PENG P. Research on collaborative filtering recommendation algorithm based on user trust and interest[D]. Njing: Nanjing University of Aeronautics and Astronautics, 2018.
- [19] ZHOU Z H, TANG W. Cluster ensemble[J]. *Knowledge Based Systems*, 2006, 19(1): 77-83.
- [20] CUI C S. Output of recommender systems based on aggregative rank[J]. *Systems Engineering-Theory & Practice*, 2013, 33(7): 1845-1851.
- [21] HU X P, YIN J. A survey of trust transfer model[J]. *Journal of Southeast University*, 2013, 15(4): 46-51, 57, 135.

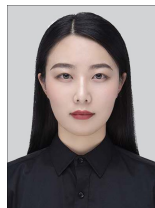


**SHAO Chao**, born in 1977, Ph.D, professor, M.S. supervisor. His main research interests include machine learning, data mining, etc.

(上接第 234 页)

## 参 考 文 献

- [1] 杨伊宁, 励建安. 应对老龄化社会: 协同构建医养融合的养老新模式[J]. *南京社会科学*, 2019(2): 73-78.
- [2] 段宝玲, 冯强, 刘德军, 等. 基于灰色 GM(1,1) 模型的城市人口规模预测——以大同市为例[J]. *山西农业大学学报(自然科学版)*, 2016, 36(1): 35-38.
- [3] 赵旭, 陈立萍, 程维虎. Logistic 回归模型在人口问题中的应用[J]. *应用概率统计*, 2015, 31(6): 662-666.
- [4] 史佳, 冀巨海. 基于 BP 神经网络的黄河中下游人口预测研究[J]. *科技管理研究*, 2014, 34(6): 245-250.
- [5] 郭雪峰, 黄健元, 王欢. 改进的灰色模型在流动人口预测中的应用[J]. *统计与决策*, 2018, 34(8): 76-79.
- [6] 蒋若凡, 姜玉梅, 李菲雅. 基于灰色 PSO-BP 人口预测模型的研究与应用[J]. *西北人口*, 2011, 32(3): 23-26.
- [7] 徐翔燕, 侯瑞环. 基于 GM(1,1)-SVM 组合模型的中长期人口预测研究[J]. *计算机科学*, 2020, 47(S1): 485-487.
- [8] 赵超, 林思铭, 许巧玲. 基于 GM-RBF 神经网络的高校建筑能耗预测[J]. *南京理工大学学报*, 2014, 38(1): 48-53.
- [9] ZHANG L, XU C, GAO Y, et al. Improved Dota2 lineup recommendation model based on a bidirectional LSTM[J]. *Tsinghua Science and Technology*, 2020, 25(6): 712-720.
- [10] WANG Q, BU S, HE Z. Achieving Predictive and Proactive Maintenance for High-Speed Railway Power Equipment with LSTM-RNN[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(10): 6509-6517.
- [11] 何为, 唐智和, 吴甬, 等. 基于 LSTM 的催化裂化装置 NO<sub>x</sub> 排放预测模型及应用[J]. *西安石油大学学报(自然科学版)*, 2020, 35(4): 108-113.
- [12] 邓丽, 邹群勇, 杨水荣. 融合 SSAE 深度特征学习和 LSTM 网络的 PM2.5 小时浓度预测[J]. *环境科学学报*, 2020, 40(9): 3422-3434.
- [13] 张军芳, 张利民, 安达. 基于灰色关联度和 BP 神经网络的全国人口预测研究[J]. *农村经济与科技*, 2019, 30(22): 257-258.
- [14] 张建海, 张祺, 许德合, 等. ARIMA-LSTM 组合模型在基于 SPI 干旱预测中的应用——以青海省为例[J]. *干旱区地理*, 2020, 43(4): 1004-1013.
- [15] 景楠, 史紫荆, 舒毓民. 基于注意力机制和 CNN-LSTM 模型的沪铜期货高频价格预测[J/OL]. *中国管理科学*. <https://doi.org/10.16381/j.cnki.issn1003-207x.2020.0342>.
- [16] 徐洪福, 袁一星, 赵洪宾. 灰色预测模型在年用水量预测中的应用[J]. *哈尔滨建筑大学学报*, 2001(4): 61-64.
- [17] 代杰杰, 宋辉, 盛戈峰, 等. 采用 LSTM 网络的电力变压器运行状态预测方法研究[J]. *高电压技术*, 2018, 44(4): 1099-1106.
- [18] HOCHREITER S, SCHMIDHUBER J. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.



**CHEN Hui-qin**, born in 1997, postgraduate. Her main research interests include land economy and policy, and larithmics.



**GUO Guan-cheng**, born in 1977, Ph.D, professor, doctoral advisor. His main research interests include land economy and management.