

# 融合信任计算的协同过滤推荐方法<sup>\*</sup>

杜永萍 黄 亮 何 明

(北京工业大学 计算机学院 北京 100124)

**摘 要** 协同过滤推荐是目前应用最为广泛的推荐策略之一,但存在数据稀疏和难扩展问题.文中在传统基于用户的协同过滤推荐算法的基础上,引入信任关系计算,利用信任关系的条件传递特性,设计并构建一个集用户声望信任和用户局部信任的混和信任网络,并将用户间评分相似度和网络中用户间信任评价度结合,为用户寻找更多基于信任因素和兴趣因素的二维相似近邻.在 Epinions 数据集上以平均绝对误差(MAE)和均方根误差(RSME)等作为实验评价指标,对该方法进行验证实验.结果表明相比传统协同过滤推荐算法,该方法在 MAE 上提高约 6.8%,最优值达到 0.7513,检验的结果也表明该方法能显著提高推荐系统性能.

**关键词** 协同过滤,信任计算,推荐系统

中图法分类号 TP 182

## Collaborative Filtration Recommendation Algorithm Based on Trust Computation

DU Yong-Ping, HUANG Liang, HE Ming

(College of Computer Science, Beijing University of Technology, Beijing 100124)

### ABSTRACT

Collaborative filtration is one of the most widely used recommendation strategies, in which data sparsity problem and expansion difficulty exist. Based on traditional user-based collaborative filtering algorithms, the trust computation is introduced into the process of recommendation. Making full use of the propagation characteristics of trust relationship under some conditions, a hybrid network composed of the user reputation-trust and the user local-trust is designed and built. And the user rating similarity is combined with trust evaluation of the hybrid network, which helps users to discover more two-dimensional similarity neighbors based on trust and interest factors. The proposed method is validated by the experiment on Epinions dataset with Mean Absolute Error (MAE) and Root Mean Square Error (RSME) as the evaluation index. The results show that compared to the traditional collaborative filtering recommendation algorithms, MAE of the proposed method increases about 6.8% and the optimal value reaches 0.7513, and the *t*-test results also show that the proposed method improves the performance significantly.

<sup>\*</sup> 国家自然科学基金项目(No. 60803086)、国家科技支撑计划子课题(No. 2013BAH21B02-01)、北京市自然科学基金项目(No. 4123091)、北京市属高等学校人才强教深化计划“中青年骨干人才培养计划”项目(No. PHR20110815)资助

收稿日期:2013-04-18;修回日期:2013-07-09

作者简介 杜永萍(通讯作者),女,1977年生,博士,副教授,主要研究方向为自然语言处理、信息检索. E-mail: y pdu@bjut.edu.cn. 黄亮,男,1988年生,硕士研究生,主要研究方向为信息检索、数据挖掘. 何明,男,1975年生,博士,副教授,主要研究方向为智能信息处理、数据挖掘.

**Key Words** Collaborative Filtration , Trust Computation , Recommendation System

## 1 引言

针对网络发展过程中的信息过载问题,推荐系统(Recommendation System)应运而生.该系统通过获取并分析用户的历史行为及信息数据,预测用户的喜爱和偏好,为用户推送感兴趣的信息.主要的推荐策略包括基于内容的推荐、基于协同过滤技术的推荐、基于知识的推荐和组合推荐等<sup>[1]</sup>.其中,协同过滤推荐技术应用最广泛,主要运用“人以类聚,物以群分”的集体智慧<sup>[2]</sup>观点为用户提供个性化服务,而服务的核心在于它的推荐能力<sup>[3]</sup>,如何使用恰当的推荐算法克服或缓解推荐系统常见的冷启动、难扩展和数据稀疏等问题,提升推荐的准确率,提高推荐信息与用户兴趣的吻合度,成为提升个性化服务质量的关键.

为取得好的推荐效果,国内外相关学者和研究人员尝试不少新的思路.例如,文献[3]提出划分聚类用户,在用户聚类结果上进行邻居选择,提高推荐准确性和实时性;文献[4]利用凸优化算法对评分矩阵空白项进行填补预测处理;文献[5]通过缺省投票、倒排用户频率等方法预测评分矩阵中的缺省值,缓解数据稀疏问题;文献[6]通过引入多指标评分概念,提出基于 Windrow-Hoff 神经网络的多指标推荐算法,提高推荐精度;文献[7]通过基于用户的推荐等级、邻域相关度和评价相似度等信息过滤推荐用户,提出一种基于偏好推荐的可靠服务,解决冷启动问题的同时也使推荐信息更为可信.

在传统协同过滤推荐算法中,大部分策略仅关注于分析用户兴趣特征,常因数据稀疏导致推荐不理想.而现实生活中,人们不仅要考虑兴趣因素,信任因素也一定程度上影响着目标用户的决策.文献[8]中,Massa 等将信任引入到推荐系统中,提出信任感知的推荐系统框架,利用用户间的信任值取代协同过滤中用户相似度做评分预测.文献[9]中,

Ziegler 等证明信任与用户的相似度有一定的相关性,即信任网内成员间相似度远大于网内成员与网外成员的相似度.文献[10]和文献[11]也分别提出针对 P2P 的网络信任模型,用于解决网络推荐过程中节点间的信任问题.文献[12]结合基于信任的推荐和基于项目的协同过滤推荐,通过随机游走预测源用户对目标项目的评分.

本文设计并构建一个集用户声望信任和局部信任的混合信任网络,用以挖掘用户间潜在在信任关系,缓解数据稀疏问题,并将网络中信任评价值与用户间评分相似度进行组合,综合考虑用户间信任因素和兴趣因素,为用户选择基于信任和兴趣的二维相似近邻,替代用评分相似度或信任相似度等单因素选择相似近邻的传统方法,以此提升系统的可靠性.

本文首先介绍传统协同过滤推荐算法,然后对信任概念和信任网络的构建过程进行介绍,接着重点介绍融合信任关系的协同过滤推荐算法,包括相似度计算、信任网络计算和混合推荐建模,最后对实验与结果进行分析,包括多种性能比较等.

## 2 传统协同过滤推荐算法

协同过滤技术广泛应用于当前电子商务个性化推荐系统中,拥有较成熟的理论与应用基础.基于与目标用户有相似兴趣偏好的其他用户对某个项目评价,判断该项目对目标用户是否有推荐价值.实现方法是通过分析目标用户和整个用户群的历史评分记录来预测目标用户对未评分项的评分,向用户推荐预测评分高的项目.从图 1 的评分预测示例中可看出,用户 1 和用户 3 的评分向量中分别缺失对第一项和第三项的评分值,系统经过评分预测模型对缺省值进行评分预测,得到用户 1 对第一项和用户 3 对第三项的评分预测值分别为 4 和 3.

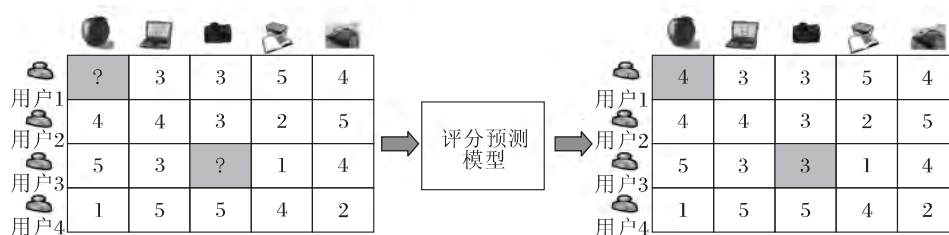


图 1 用户-项目评分矩阵及评分预测示例

Fig. 1 User-item rating matrix and prediction sample

协同过滤推荐算法的实现主要由 3 个模块组成:特征选取、相似度计算和评分预测。图 2 为推荐系统的整体结构。首先,系统对目标用户进行特征选取,选择能表征目标用户与其他用户间关系的特征或属性;然后,对选择的特征属性构建相似度计算模型,并将用户间相似度值存储在关系数据库中;最

后,进行评分预测,从关系数据库中寻找若干与目标用户相似度较高的相似近邻,从用户偏好数据库中提取相似近邻的偏好评分记录,去除目标用户历史评分项目后,将余下项目作为候选推荐项目集,利用评分预测模型对候选集中项目进行推荐评分计算,选取推荐评分较高的项目为目标用户实行项目推荐。

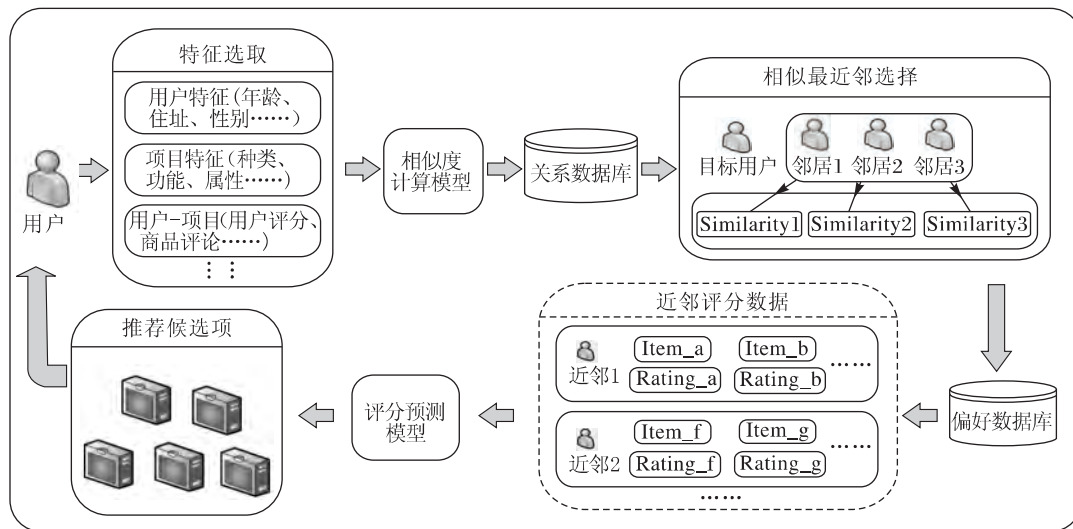


图 2 协同过滤推荐系统结构

Fig. 2 Structure of collaborative filtering recommendation system

### 3 基于信任计算的信任网络构建

本文将信任引入协同过滤推荐算法中,认为信任是系统中某用户在综合了解其他用户历史行为后,对他们推荐行为的一种认可,是对这些推荐信息的真实性、安全性及其实用价值的一种主观肯定,将随着时间等因素变化而改变。

信任网络构建实际是指通过获取并分析用户间的历史行为数据,挖掘用户影响力及用户间信任关系(直接信任和间接信任),并结合图论的相关知识构建出一个加权有向图,构建过程包括信任网络的初始化和局部信任网络的扩展两部分。

#### 3.1 信任网络初始化

结合图论相关知识,将用户视为图中节点,若用户间存在信任关系,则图中相应节点间必存在一条有向边,边上的权值代表用户间的信任值,那么由节点集合  $V$  和有向边集合  $E$  组成的有向图  $G = (V, E)$  即为基于信任关系的信任网络。在有向图  $G$  的基础上,利用用户的信任度作为声望信任的计算参数,进行用户声望信任( $Auth$ )的归一化计算,计算公式如下:

$$Auth(A) = \frac{Indg(G, A) - \min(Indg(G, ))}{\max(Indg(G, )) - \min(Indg(G, ))}$$

其中  $Indg(G, A)$  表示图  $G$  中用户  $A$  的信任入度,  $\max(Indg(G, ))$  和  $\min(Indg(G, ))$  分别表示图  $G$  中用户最大信任入度和最小信任入度。该公式得到的声望信任值越大表明该用户影响力越大,越容易得到其他用户的信任。

#### 3.2 局部信任网络扩展

在对信任网络进行初始化后,利用相应的信任关系传递规则对网络中用户的局部信任网络进行扩展,获取用户间接信任关系。如图 3 所示,图中圆形节点(如  $A$  节点)代表用户,且每个用户都拥有一个基于直接信任关系的声望信任值(如  $A$  用户节点的  $Auth(A) = 0.2$ )。实线有向边表明该节点对之间拥有直接信任关系,虚线有向边表明节点用户间的间接信任值。

以图中用户  $A$  节点的局部信任网络扩展为例,从起始节点  $A$  开始,广度搜索与节点  $A$  有直接信任关系(网络路径长度为 1)的节点  $\{B, C, D, F\}$  的网络邻居,获得网络路径长度为 2 的间接网络邻居  $\{E, G, H, I\}$ ,然后依次遍历  $E, G, H, I$  节点的网络邻居,形成距用户节点  $A$  网络路径长度为 3 的网络邻居。

集 依次迭代,直到达到系统预定的局部信任网络最大规模或迭代无变化,至此可找出与  $A$  存在广义信任关系的邻居集合,从而形成用户  $A$  的局部信任网络。下文将针对用户  $A$  与网络邻居间的间接信任值则采用局部信任和声望信任组合计算的方法进行介绍。

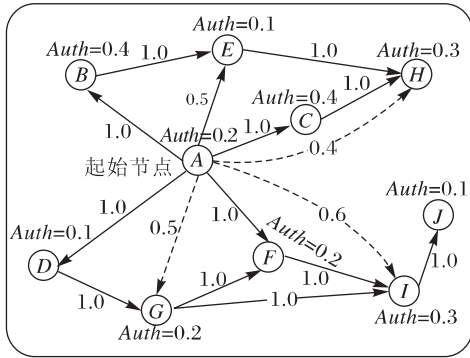


图3 信任网络示例

Fig.3 Trust network illustration

#### 4 融合信任关系的协同过滤推荐方法

本文将信任关系融合到协同过滤推荐过程中,系统结构如图4所示,在传统协同过滤的基础上利用用户间的直接信任关系和信任关系传递规则,构建并扩展一个集用户声望信任和局部信任的混和信任网络。组合用户间兴趣相似度 (Similarity)

与混合信任网络中信任评价度 (Trust) 为组合权重 (Join\_Sim), 以此选择目标用户基于兴趣和信任因素的二维相似近邻。融合信任关系的协同过滤算法 (Combined Recommendation ( $R, T, \beta$ )) 过程如下。

算法 融合信任关系的协同过滤推荐算法

输入  $R$  (用户-项目评分矩阵)  $T$  (用户-用户信任矩阵)  $\beta$  (最近邻选择阈值)

输出  $P$  (预测评分矩阵)

Begin

$S^1 \leftarrow \text{User\_Similarity\_Computing\_Model}(R)$

//用户间相似度计算模型

$S^2 \leftarrow \text{User\_Trust\_Network\_Computing\_Model}(T)$

//信任网络计算模型

For  $a \leftarrow 1$  to  $n$  do

For  $i \leftarrow 1$  to  $m$  do

$P_{a,i} \leftarrow \text{combined\_predict\_model}(S^1, S^2, \beta)$  //组合推荐模型

End

Return  $P$

参数描述:  $\beta$  为预先设置的用户最近邻选择阈值,只有相似度大于  $\beta$  值的用户才能成为邻居;  $S^1$  中存储基于用户的相似度矩阵;  $S^2$  中存储基于信任挖掘的信任网络矩阵;  $P$  为针对用户的预测评分矩阵,  $P_{a,i}$  表示算法给出的用户  $a$  对项目  $i$  的预测评分。下面分别介绍算法中提到的“用户间相似度计算模型”、“信任网络计算模型”和“组合推荐模型”。

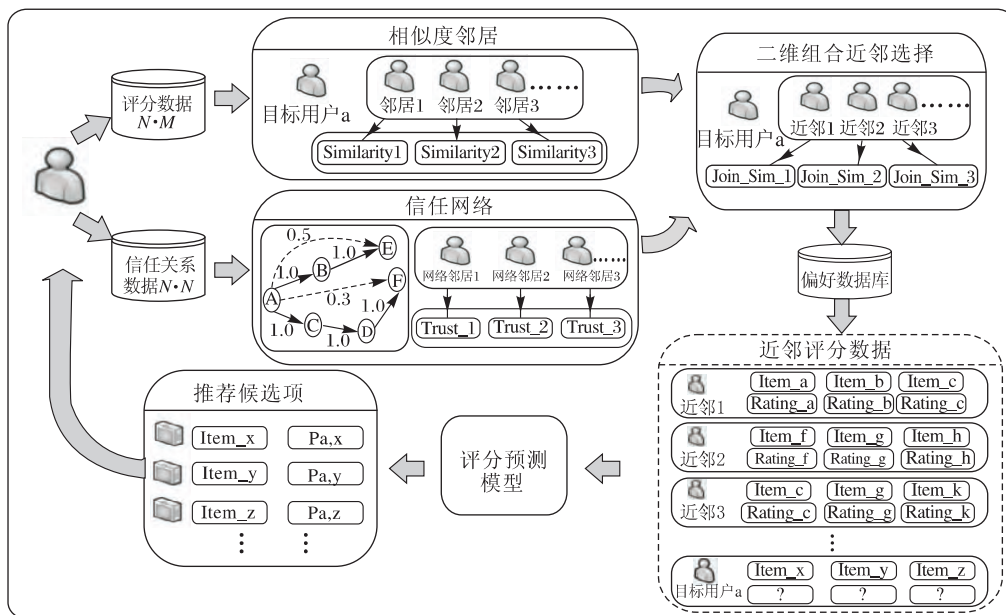


图4 融合信任关系的协同过滤推荐算法框架

Fig.4 Framework of collaborative filtering recommendation algorithm based on trust relationship

#### 4.1 相似度计算模型

本文利用用户-项目评分矩阵计算用户间基于偏好立场的兴趣相似度 (User Preferences Position Similarity, UPPS). 首先获取用户间评分项目的交集, 并设定表征用户对项目偏好立场的阈值  $R$  (如  $R=2$ ). 若评分值  $\geq R$ , 表明用户对项目较感兴趣, 否则表示用户对项目不感兴趣. 若两个用户有共同评分项目集且相同偏好倾向 (包括感兴趣和不感兴趣) 所占的比例越大, 则用户兴趣越相似. 用户间基于偏好立场的兴趣相似度计算公式如下:

$Similarity(a, b) =$

$$\left( \frac{|I_a \cap I_b|^{\geq R}}{|I_a|^{\geq R}} + \frac{|I_a \cap I_b|^{< R}}{|I_a|^{< R}} \right) \cdot \left( \frac{|I_a \cap I_b|^{\geq R}}{|I_b|^{\geq R}} + \frac{|I_a \cap I_b|^{< R}}{|I_b|^{< R}} \right) \quad a \neq b,$$

其中, 参数  $R$  表明用户偏好立场的阈值. 如果系统中的评分等级为 1 至 5 之间的正整数, 则  $R$  取值也在 1 至 5 之间.  $|I_a|^{\geq R}$  表示用户  $a$  的评分记录中评分值大于或等于  $R$  的项目集合规模.  $|I_a \cap I_b|^{\geq R}$  表示用户  $a$  与  $b$  评分交集中评分值大于或等于  $R$  的项目集合规模.

#### 4.2 信任网络计算模型

现实生活中, 人们不仅信任自己的朋友, 而且可能在已有信任朋友的作用下与朋友的朋友建立起新的信任关系. 将这一现实场景重现到信任网络中: 若用户  $A$  信任用户  $B$ , 用户  $B$  信任用户  $C$ , 而用户  $A$ 、 $C$  间无直接信任关系. 在  $A-B-C$  可达路径条件下, 用户  $A$ 、 $C$  间可能存在新的信任关系, 即信任具有条件可传递性. 本文将利用这一特性对局部信任网络进行扩展. 在获取用户间的直接信任关系后, 依据相应的信任关系传递规则, 获取用户局部信任网络中没有直接信任关系但存在可达路径用户间的信任关系.

根据社会学中有关信息传播的研究可知信息在社会网络中传播的路径越长, 可信度就越低. 借鉴这一研究理论并结合前面的现实场景, 本文在信任关系传递的过程中做出相应假设: 信任关系在传递过程中会出现一定程度的衰减, 且传递路径越长, 衰减

就越严重. 根据这个假设, 制定信任关系在传递过程中必须遵守的两个传递规则.

1) 规则 1:

$$\forall (DirectTrust(a, b), DirectTrust(b, c) \in [0, 1]) \Rightarrow IndirectTrust(a, c) \in [0, 1].$$

2) 规则 2:

$$\forall (DirectTrust(a, b), DirectTrust(b, c) \in [0, 1]) \Rightarrow IndirectTrust(a, c) < \min(DirectTrust(a, b), DirectTrust(b, c)).$$

$DirectTrust(a, b)$  表示用户  $a$  与  $b$  之间的直接信任值.  $IndirectTrust(a, c)$  表示用户  $a$  与用户  $c$  之间的间接信任值. 规则 1 作为值域限制, 定义信任值取值只能在  $[0, 1]$  之间; 规则 2 表明信任的传递通常随路径长度的增加而逐渐衰减.

##### 4.2.1 局部信任网络扩展

鉴于规则 1 和规则 2 的要求, 信任关系传递过程中必须满足约束条件: 间接信任值取值区间为  $[0, 1]$ ; 信任关系传递过程中信任值会随传递路径的增长而衰减. 因此, 本文采用能同时满足上述两个条件的乘法函数来计算间接信任关系. 局部信任网络扩展过程中信任关系的传递与计算过程如下.

1) 选取存在可达路径的网络邻居用户.

将信任网络中某用户  $a$  作为起始用户, 从该用户节点出发, 在信任网络中进行广度迭代搜索, 寻找距用户  $a$  路径长度分别为  $PathLen = \{1, 2, \dots, N\}$  的其他用户作为网络邻居用户. 该邻居用户满足与用户  $a$  直接相邻或间接相邻 (在  $PathLen$  路径长度内路径可达). 但随着扩展路径长度  $PathLen$  的增加, 可能会产生噪音数据, 为避免这种情况, 设置最大扩展路径长度  $\max(PathLen) = 6$ .

2) 计算单路径用户间的间接信任值.

若网络中起始用户  $a$  与邻居用户  $c$  存在可达网络路径, 如  $Path = (a, t_1, t_2, \dots, t_m, c)$ , 其中  $t_1, t_2, \dots$  都是该可达路径上的中间用户, 则该路径上用户  $a$  与用户  $c$  之间的间接信任值和该路径的长度成反比, 可利用如下公式进行计算:

$$IndirectTrust(a, c) = \frac{DirectTrust(a, t_1) \cdot DirectTrust(t_1, t_2) \cdot \dots \cdot DirectTrust(t_m, c)}{PathLen(a, c)},$$

其中  $PathLen(a, c)$  表示可达路径的长度.

3) 计算多路径用户间的间接信任值.

若起始用户与网络邻居用户间存在多条网络可达路径, 则可采用如下公式来计算起始用户到网络邻居用户的间接信任值:

$$IndirectTrust(a, c)' =$$

$$\frac{1}{|Path(a, c)|} \sum_{Path(a, c)} IndirectTrust(a, c),$$

其中  $Path(a, c)$  表示用户  $a$  与  $c$  的可达路径,  $|Path(a, c)|$  表示可达路径的数目.



#### 4.2.2 混合信任网络构建

网络中用户间信任关系需要考虑用户间的局部

信任,也需考虑邻居用户在网络中的声望信任.因此,本文设计用户间信任关系值  $W_{trust}$  的计算公式如下:

$$W_{trust}(a, c) = \begin{cases} \xi \cdot IndirectTrust(a, c) + (1 - \xi) \cdot Auth(c), & \text{用户 } a, c \text{ 间接相邻} \\ \xi \cdot DirectTrust(a, c) + (1 - \xi) \cdot Auth(c), & \text{用户 } a, c \text{ 直接相邻} \end{cases}$$

其中,  $W_{trust}(a, c)$  为信任网络中用户  $a$  与用户  $c$  间的信任关系值(直接信任关系或间接信任关系)  $Auth(c)$  为用户  $c$  的声望信任值  $\xi$  为 0 到 1 间的调节系数.

在信任传播过程中,以上信任关系计算无法区别网络中不同兴趣偏好用户间的信任关系,而拥有共同兴趣的用户间信任关系应该比普通信任关系更为密切.因此,在用户间信任关系值  $W_{trust}$  计算公式的基础上增加对用户间兴趣信任的描述项,使得与目标用户兴趣越相近的用户其信任值更高.以此来区分信任关系在不同兴趣偏好用户间的传递,信任评价度  $Trust$  计算公式如下:

$$Trust(a, c) = \mu \cdot W_{trust}(a, c) + (1 - \mu) \cdot \frac{|I_a \cap I_b| \geq R}{|I_a \cap I_b|}.$$

用户  $a$  与用户  $c$  间的信任评价度  $Trust(a, c)$  为用户  $a$  与  $c$  的广义信任关系权值  $W_{trust}$  和兴趣信任描述项  $|I_a \cap I_b| \geq R / |I_a \cap I_b|$  的加权线性组合.  $|I_a \cap I_b|$  表示用户  $a$  与  $c$  的共同评分项的规模,  $|I_a \cap I_b| \geq R$  表示用户  $a$  和  $c$  共同评分项中评分值大于等于  $R$  项的规模,即用户  $a$  与用户  $c$  都感兴趣的项,以此来表示对方兴趣品味的信任值.在共同评分项中,感兴趣的项所占比例越大,说明用户间有更多相似的兴趣品味,彼此间更容易建立信任关系.

信任关系在用户间的传播除了要满足路径可达条件外,还须满足一定的阈值条件,设定信任传播的下限值  $\sigma$ ,则系统中只有信任评价度值满足下限值  $\sigma$  的用户间才能进行信任传播.在下限阈值控制下,信任评价度取值如下所示:

$$Trust(a, c) = \begin{cases} 0, & Trust(a, c) < \sigma \\ Trust(a, c), & Trust(a, c) \geq \sigma \end{cases}.$$

#### 4.3 组合推荐模型

用户间的相似关系受兴趣和信任两方面因素的影响和制约,为了在推荐过程中体现两项因素在相似近邻选取中的共同作用,进行评分预测前,需要对相似度和信任评价度进行调和,利用统计学中常用的加权调和平均数计算方法<sup>[13]</sup>进行计算,调和函数如下所示:

$$\bar{H} = \frac{\sum_{i=1}^k m_i}{\sum_{i=1}^k \frac{m_i}{x_i}} = \frac{m_1 + m_2 + \dots + m_k}{\frac{m_1}{x_1} + \frac{m_2}{x_2} + \dots + \frac{m_k}{x_k}},$$

其中  $\bar{H}$  为加权调和平均数  $m_i$  为权衡变量  $x_i$  对调和平均数  $\bar{H}$  贡献大小的权值系数,表示变量  $x_i$  的重要性.考虑到兴趣因素和信任因素在推荐过程中起着同等重要的作用,因此,两因素权值采取等值分配策略,用户  $a$  与  $c$  之间综合相似度调和平均值计算公式如下:

$$Join\_sim(a, c) = \frac{2 \times Similarity(a, c) \cdot Trust(a, c)}{Similarity(a, c) + Trust(a, c)}.$$

对系统中目标用户与相似邻居进行综合相似度计算后,选出目标范围的二维相似最近邻,将近邻评分数据作为输入提供给评分预测模型,计算推荐候选项及推荐值.推荐值的中心加权预测公式如下:

$$P_{a,i} = \bar{r}_a + \frac{\sum_{c \in S_a} (R_{c,i} - \bar{r}_c) \cdot Join\_sim(a, c)}{\sum_{c \in S_a} |Join\_sim(a, c)|},$$

其中  $P_{a,i}$  表示用户  $a$  对项目  $i$  的预测评分值  $S_a$  为用户  $a$  的最近邻居集合;  $R_{c,i}$  为用户  $c$  对项目  $i$  的评分值  $\bar{r}_a$  和  $\bar{r}_c$  分别表示用户  $a$  和用户  $c$  所有评分记录的平均评分.

## 5 实验与结果分析

### 5.1 实验数据集

本文采用 Epinions 真实数据集 (<http://www.Epinions.com>) 对本文提出的算法进行验证实验.它是一个在线服务网站,用户可以在网站上对物品进行评分,也可以将某人加入到自己的信任列表中. Epinions 数据集包含用户对项目的评分信息和用户间的直接信任关系信息,该实验数据集的统计信息如表 1 所示.

表 1 Epinions 实验数据集统计信息  
Table 1 Statistics of Epinions dataset

数据类型	用户数量	项目数量	评分/信任等级	评分数量	信任数量
评分数据	40163	139738	{1, 2, 3, 4, 5}	664824	—
信任数据	40163	—	{0, 1}	—	442979

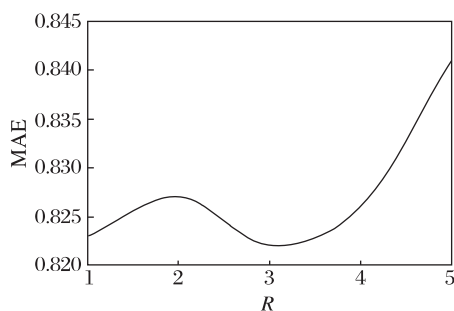
由于此数据集极其稀疏,因此本文实验数据集是 Epinions 数据集的一个相对稠密的子集.实验过

程中,首先选取一定规模的初始用户集,过滤掉该用户集中评分数量少于 20 个的用户,然后采用留一交叉检验法,以 20% 的比例抽取每个用户的评分数据作为评测数据集进行实验。

## 5.2 评价标准

推荐预测准确度是评价推荐系统推荐质量的一个重要度量标准,该指标刻画了用户针对项目满意程度的预测值与实际评分值间的匹配程度。通常采用平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Square Mean Error, RSME) 作为推荐准确性的评价指标。其中 MAE 定义如下:

$$MAE = \frac{\sum_{(a,i) \in X} |r_{a,i} - p_{a,i}|}{|X|}$$



(a) 参数 R 对 MAE 的影响

(a) Influence of parameter R to MAE

RSME 定义如下:

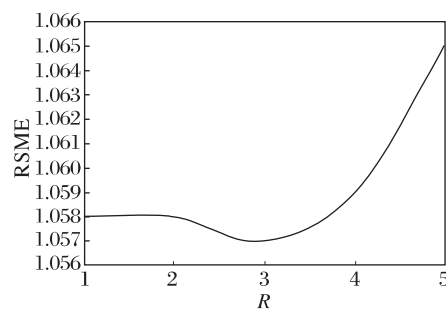
$$RSME = \sqrt{\frac{\sum_{(a,i) \in X} (r_{a,i} - p_{a,i})^2}{|X|}}$$

其中  $X$  表示评测数据集,  $r_{a,i}$  表示评测数据集中用户  $a$  针对项目的实际评分,而  $p_{a,i}$  表示推荐系统给出的用户  $a$  对项目  $i$  的预测评分值。

## 5.3 实验结果及分析

### 5.3.1 参数 R 对系统性能的影响

为检测用户偏好立场相似度 (UPPS) 计算公式在参数  $R$  取不同值时对系统性能的影响,本文设计采用基于 UPPS 的协同过滤方法进行评分预测。实验结果如图 5 所示。



(b) 参数 R 对 RSME 的影响

(b) Influence of parameter R to RSME

图 5 参数 R 对评价指标的影响

Fig. 5 Influence of parameter R to evaluation indexes

由图 5 可看出,UPPS 方法在  $R=3$  时,评价指标 MAE 和 RSME 都能取得较好结果,分别为 0.822 2 和 1.056 8,说明在 Epinions 数据集上,基于用户偏好立场的相似度计算方法不仅可以较为准确的估计用户之间的相似度,而且  $R=3$  这个值也表明在 1 到 5 的评分区间中,对多数用户来说,评分值 3 是界定用户对项目兴趣喜好程度的一个阈值,评分值高于阈值的评分记录反映出用户对项目较感兴趣,反之则表示用户对项目不感兴趣。

### 5.3.2 不同局部信任网络规模和近邻阈值 $\beta$ 对实验的影响

信任网络扩展过程中通过控制扩展路径长度可

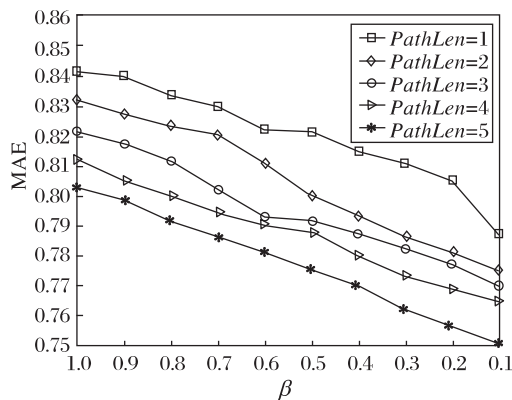
得到不同规模的局部信任网络,通过调整近邻选择阈值  $\beta$  可以获取不同规模的二维相似近邻。因此,实验设计如下:1) 随机选取 1 000 个用户,构建混和信任网络,扩展路径长度从  $PathLen=1$  (即路径长度为 1,直接相邻) 逐渐变化到  $PathLen=5$  (即用户间存在可达路径,且路径长度为 5),观察扩展后用户的平均信任关系数值变化;2) 二维组合近邻选取阈值  $\beta$  从 1.0 变化至 0.1,观察实验评价指标的变化。设计以上实验,分别用来检测不同局部信任网络规模和近邻阈值  $\beta$  取值对本文所提基于偏好立场相似度的协同过滤算法 (UPPS + Trust) 的影响。实验结果如表 2 和图 6 所示。

表 2 不同扩展路径长度下平均信任关系

Table 2 Average trust relation on different path length

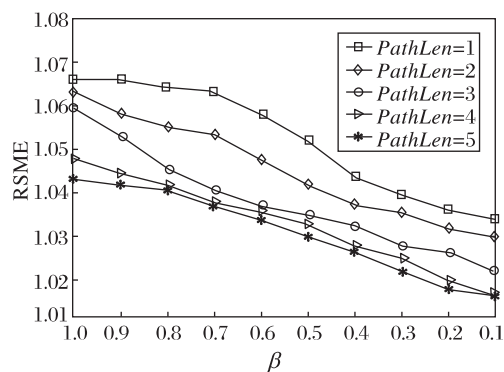
	$PathLen=1$	$PathLen=2$	$PathLen=3$	$PathLen=4$	$PathLen=5$
平均信任关系 (取整)	18	47	142	164	171

从表 2 可看出,  $PathLen$  增长过程中, 用户的平均信任关系也在不断增加, 越来越多新的信任关系被发现, 有效缓解关系数据稀疏的问题, 在  $PathLen = 3$  时增长最快, 在  $PathLen = 4$  和 5 时, 新的信任关系数增长缓慢, 说明大部分潜在信任关系已经被获取。



(a) 对 MAE 评价指标的影响

(a) Influence to MAE value



(b) 对 RSME 评价指标的影响

(b) Influence to RSME value

图 6 不同局部信任网络规模和阈值  $\beta$  对评价指标值影响  
Fig. 6 Influence of different scale of local trust network and  $\beta$ -value to evaluation indexes

从图 6 可看出, 曲线的走势受二维组合近邻阈值  $\beta$  取值和网络扩展路径长度  $PathLen$  两方面因素影响. 在阈值  $\beta$  从 1.0 变化到 0.1 的过程中, 各项评价指标值都呈现出逐渐减小的趋势 (误差变小, 准确率提高). 原因可能是降低近邻阈值  $\beta$  的取值可以为用户匹配到更多的二维相似近邻, 虽然也可能引起一定的噪声问题, 但从总体上看还是保持积极影响. 对于网络扩展路径长度  $PathLen$  对评价的影响, 综合表 2 和图 6 可看出, 在  $PathLen = 1$  (只包含直接信任关系) 时, 评测指标值较高, 因为此时网

络邻居数量较少, 稀疏的信任关系影响结果, 而当  $PathLen$  从 1 到 5 变化过程中, 评价指标值逐渐降低 (误差变小, 准确率提高), 表明网络扩展路径长度的增加使得更多信任关系被发现, 缓解了关系数据稀疏影响, 产生积极影响.

### 5.3.3 不同推荐策略的性能比较

对比不同推荐策略的实验性能, 包括采用 Pearson 系数的传统协同过滤推荐策略 (Pearson CF)、基于用户偏好立场兴趣相似度的协同过滤推荐策略 (UPPS)、基于信任关系的推荐策略 (Trust)、融合信任关系的传统协同过滤推荐策略 (Pearson + Trust)、融合信任因素和兴趣因素的 UPPS 协同过滤推荐策略 (UPPS + Trust) 实验结果如表 3 所示. 利用表 3 实验结果进行配对  $t$  检验, 检测信任的引入对实验性能是否产生积极影响, 配对  $t$  检验结果如表 4 所示.

表 3 不同相似度计算方法的 MAE 对比

Table 3 MAE comparison of different similarity measurement methods

$\beta$	Pearson	UPPS	Trust	Pearson + Trust	UPPS + Trust
1.0	0.8457	0.8390	0.8288	0.8175	0.8034
0.9	0.8432	0.8337	0.8251	0.8134	0.7986
0.8	0.8423	0.8367	0.8233	0.8109	0.7917
0.7	0.8395	0.8346	0.8214	0.8096	0.7861
0.6	0.8373	0.8163	0.8195	0.8077	0.7813
0.5	0.8356	0.8143	0.8169	0.8061	0.7758
0.4	0.8313	0.8119	0.8135	0.8037	0.7701
0.3	0.8263	0.8086	0.8113	0.8014	0.7623
0.2	0.8242	0.8045	0.8107	0.7995	0.7565
0.1	0.8217	0.7994	0.8101	0.7987	0.7513

表 4 不同推荐策略的 MAE 配对  $t$  检验

Table 4 Paired  $t$ -test for MAE of different recommendation strategies

方法	$df$	$t$	$p$ -value
Pearson vs. Pearson + Trust	18	8.4230	5.8383e-08
UPPS vs. UPPS + Trust	18	5.7960	8.5802e-06

从表 3 和表 4 中可看出, 本文提出的融合信任因素和兴趣因素的 UPPS 协同过滤推荐策略 (UPPS + Trust) 取得较好的性能, MAE 最优值达到 0.7513, 说明该方法对提高推荐系统性能产生积极影响. 融合信任关系的双因素协同过滤推荐策略 (包括 Pearson + Trust 和 UPPS + Trust) 相比单因素推荐策略 (如 Pearson、UPPS、Trust) 在整体上也取得较好的性能, 说明利用双因素进行相似近邻选择比



单因素更有效,增强系统对用户和产品规模变化的适应能力,提高系统的可扩展性。

除此之外,将本文推荐策略和文献[12]中所提推荐策略进行性能对比,从UPPS+Trust的RSME评价指标变化(如图6所示)可看出,本文推荐策略RSME最优值1.015,优于文献[12]的最优值1.192(对冷启动用户)和1.077(对所有用户),说明本文融合信任关系计算的协同过滤推荐方法优于文献[12]采用的随机游走方法,在综合考虑信任和兴趣双重因素的基础上进行二维相似近邻选取,具有更高的可靠性。而文献[12]将基于项目的协同过滤算法加入到基于信任关系随机游走的推荐系统中,没有考虑用户间兴趣因素的变化。

## 6 结 束 语

本文将信任关系作为影响推荐的一个重要因素,利用信任关系的条件可传递性,设计并构造一个混和信任网络,并在此基础上为目标用户选取基于兴趣和信任双重因素的二维相似近邻。实验结果表明,本文提出的融合信任关系的协同过滤推荐策略为目标用户获取到新的潜在信任关系,有效缓解用户间关系数据稀疏的问题,且利用双因素代替单因素选取相似近邻的方法,为目标用户匹配到更多的二维相似邻居,提高系统的可扩展性,使系统更容易适应商品和用户激增变化的影响,提高系统的可靠性。

## 参 考 文 献

- [1] Xu H L, Wu X, Li X D, *et al.* Comparison Study of Internet Recommendation System. *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. *软件学报*, 2009, 20(2): 350 - 362)
- [2] Wang L C, Meng X W, Zhang Y J. Context-Aware Recommender Systems. *Journal of Software*, 2012, 23(1): 1 - 20 (in Chinese)  
(王立才, 孟祥武, 张玉洁. 上下文感知推荐系统. *软件学报*, 2012, 23(1): 1 - 20)
- [3] Wu H C, Wang X J, Cheng Y, *et al.* Advanced Recommendation Based on Collaborative Filtering and Partition Clustering. *Journal of Computer Research and Development*, 2011, 48 (Supplement 2): 205 - 212 (in Chinese)  
(吴泓辰, 王新军, 成勇, 等. 基于协同过滤与划分聚类的改进推荐算法. *计算机研究与发展*, 2011, 48(增刊2): 205 - 212)
- [4] Candès E J, Recht B. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 2009, 9(6): 717 - 772
- [5] Breese J S, Hecherman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering // *Proc of the 14th Conference on Uncertainty in Artificial Intelligence*. Madison, USA, 1998: 43 - 52
- [6] Zhang F Z, Chang J F, Wang D. Multi-criteria Recommendation Algorithm Based on Widrow-Hoff Neural Network. *Pattern Recognition and Artificial Intelligence*, 2011, 24(2): 233 - 242 (in Chinese)  
(张付志, 常俊凤, 王栋. 基于 Widrow-Hoff 神经网络的多指标推荐算法. *模式识别与人工智能*, 2011, 24(2): 233 - 242)
- [7] Zhu R, Wang H M, Feng D W. Trustworthy Services Selection Based on Preference Recommendation. *Journal of Software*, 2011, 22(5): 852 - 864 (in Chinese)  
(朱锐, 王怀民, 冯大为. 基于偏好推荐的可信服务选择. *软件学报*, 2011, 22(5): 852 - 864)
- [8] Massa P, Avesani P. Trust-Aware Collaborative Filtering for Recommender Systems // *Proc of the OTM Confederated International Conferences: CoopIS, DOA, and ODBASE*. Agia Napa, Cyprus, 2004: 492 - 508
- [9] Ziegler C N, Lausen G. Analyzing Correlation between Trust and User Similarity in Online Communities // *Proc of the 2nd International Conference on Trust Management*. Oxford, UK, 2004: 251 - 265
- [10] Wei F, Li J, Hu J M. New Trust Model Based on Preference Recommendation in P2P Network. *Application Research of Computers*, 2010, 27(6): 2271 - 2272, 2279 (in Chinese)  
(魏锋, 李杰, 胡江明. 一种基于优先度推荐的新型 P2P 网络信任模型. *计算机应用研究*, 2010, 27(6): 2271 - 2272, 2279)
- [11] Yu Z, Shen G C, Liu B W, *et al.* METrust: A Trust Model in P2P Networks. *Acta Electronica Sinica*, 2010, 38(11): 2600 - 2605 (in Chinese)  
(于真, 申贵成, 刘丙午, 等. 一种 P2P 网络信任模型 METrust. *电子学报*, 2010, 38(11): 2600 - 2605)
- [12] Jamali M, Ester M. TrustWalker: A Random Walk Model for Combining Trust-Based and Item-Based Recommendation // *Proc of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Paris, France, 2009: 397 - 406
- [13] Jiao H H, Li Y, Cheng Q. Statistical Theory. Chengdu, China: Southwestern University of Finance and Economics Press, 2004 (in Chinese)  
(焦洪浩, 李勇, 陈琴. 统计学原理. 成都: 西南财经大学出版社, 2004)