

## 社交网络环境下基于信任的推荐算法\*

陈婷<sup>2</sup>, 朱青<sup>1,2</sup>, 周梦溪<sup>2</sup>, 王珊<sup>1,2</sup>

<sup>1</sup>(数据工程与知识工程教育部重点实验室(中国人民大学), 北京 100872)

<sup>2</sup>(中国人民大学 信息学院, 北京 100872)

通讯作者: 朱青, E-mail: zq@ruc.edu.cn



**摘要:** 现有的基于信任的推荐算法通常假设用户是单一和同质的,没有充分挖掘信任关系信息,且相似关系和信任关系的融合缺乏高效的模型,极大地影响了推荐的准确性和可靠性。提出一种基于信任的推荐算法。首先,结合全局信任和局部信任,并利用信任的传播性质对信任关系进行建模;然后,设置推荐权重,综合考虑相似度和信任度来构建用户间的偏好关系,筛选出邻居;最后,将基于记忆的协同过滤思想和社交网络的信任关系融入概率矩阵分解模型,同时使用自适应权重动态决定各部分的影响程度,形成高效、统一的可信推荐模型 Trust-PMF。该算法在 FilmTrust, Epinions 这两个数据集上与相关算法做了对比验证,结果证实了该算法的高效性。

**关键词:** 社会网络;信任;概率矩阵因子分解;推荐系统

**中图法分类号:** TP311

中文引用格式: 陈婷,朱青,周梦溪,王珊. 社交网络环境下基于信任的推荐算法. 软件学报, 2017, 28(3): 721-731. <http://www.jos.org.cn/1000-9825/5159.htm>

英文引用格式: Chen T, Zhu Q, Zhou MX, Wang S. Trust-Based recommendation algorithm in social network. Ruan Jian Xue Bao/Journal of Software, 2017, 28(3): 721-731 (in Chinese). <http://www.jos.org.cn/1000-9825/5159.htm>

## Trust-Based Recommendation Algorithm in Social Network

CHEN Ting<sup>2</sup>, ZHU Qing<sup>1,2</sup>, ZHOU Meng-Xi<sup>2</sup>, WANG Shan<sup>1,2</sup>

<sup>1</sup>(Key Laboratory for Data and Knowledge Engineering, Ministry of Education (Renmin University of China), Beijing 100872, China)

<sup>2</sup>(Information School, Renmin University of China, Beijing 100872, China)

**Abstract:** The existing trust-based recommendation algorithms usually assume that users are homogeneous, and therefore can't fully mine the trust relationship information. Moreover, the lack of efficient model for integrating similar relationship and trust relationship greatly affects the accuracy and reliability of those models. To solve the issue, this paper first proposes a trust-based recommendation algorithm called Trust-PMF. It combines similarity with trust to build user's preference and selects the target user's neighbors. Then, the probability matrix factorization model is extended by integrating memory-based idea and trust information, and a dynamic adaptive weight is used to determine the degree of influence of each part to form a unified and efficient Trust-PMF model. Finally, experiment results on Filmtrust and Epinions data sets are presented to demonstrate that the proposed method outperforms the state-of-the-art methods.

**Key words:** social network; trust; PMF; recommender system

社交网络信息是推荐系统的重要信息来源。随着 Facebook、Linkedin、Twitter、instagram、微博、微信等基于 Web 2.0 技术的社交网络爆增长,目前,一些电子商务网站已经基于社交网络构建,如阿里巴巴入股新浪,借助

\* 基金项目: 国家自然科学基金(61070053)

Foundation item: National Natural Science Foundation of China (61070053)

收稿时间: 2016-07-25; 修改时间: 2016-09-14; 采用时间: 2016-11-11; jos 在线出版时间: 2016-11-29

CNKI 网络优先出版: 2016-11-29 13:35:01, <http://www.cnki.net/kcms/detail/11.2560.TP.20161129.1335.006.html>

微博强化社交化的属性,加速实现社会化电商.人人网、Epinions、微博等发挥产品本身的优势将信任关系作为推荐系统的一种重要维度.事实表明:消费者更愿意相信朋友或者有影响力的人的推荐,而不是缺乏说服力的广告.信任作为人际关系的核心概念,将直接影响用户的决策过程,成为社会化推荐的研究重点.

大数据环境下的信任推荐<sup>[1]</sup>系统是传统推荐的延伸,用户面临复杂信息环境、多维数据特征和准确推荐需求.如何准确提取和预测在大数据环境下的用户偏好,生成高准确度的推荐,是大数据给推荐系统带来的机遇与挑战,具体表现在:数据产生的速度更快、内容采样渠道更多、数据高维稀疏、数据内容丰富但结构复杂、多源数据融合的噪声和冗余.另一方面,商品个数和种类快速增长,导致顾客需要花费大量的时间才能找到自己想买的商品.浏览大量无关信息的过程,无疑会使消费者流失加剧.个性化推荐系统精准且快速地协助用户决策显得越发重要.个性化推荐的方式一般细分为基于内容、基于知识、协同过滤、混合过滤推荐以及日益流行的基于社会网络推荐,但是存在冷启动、数据稀疏性、评分数据的伪造、推荐精度等问题.

社会网络环境下基于信任的推荐方法研究取得了一定的成果,主要有链接预测和矩阵分解方法.如:EigenTrust 模型是 Kamvar 等人<sup>[2]</sup>提出的一种点对点网络的信誉管理算法;TidalTrust 由 Golbeck 等人<sup>[3]</sup>提出,算法在预测评分时参考了源节点信任的所有邻居对项目的评分;MoleTrust 由 Massa 等人<sup>[4,5]</sup>提出,算法基于 TidalTrust 改进,主要靠经验,因此准确度受取值的影响较大,模型不够稳定;随机游走(TrustWalker)是 Jamali 等人<sup>[6]</sup>提出的推荐策略,随着漫游步数的增多,算法能够在预测过程中计算可信度.社交网络环境下基于信任的推荐研究还有从系统层面、语义层面<sup>[7,8]</sup>等,但是模型与我们的 Trust-PMF 混合推荐算法研究层面略有不同,值得借鉴.

矩阵因子分解框架上扩展的信任推荐.如:由 Ma 等人提出的 SoRec<sup>[9]</sup>方法将信任网络结构和用户对项目的评分信息,融入概率矩阵因子模型;Ma 等人<sup>[10]</sup>又提出另外一种基于信任的推荐算法 RSTE;Jamali 和 Ester<sup>[11]</sup>提出的 SocialMF 方法弥补了 RSTE 的不足;郭磊、马军等人<sup>[12]</sup>提出了 StrengthMF 和 InfluenceMF 算法,算法对用户的信任关系和兴趣偏好通过共享的特征空间建模,基于 SoRec 方法提出社会化推荐算法 PMFUI<sup>[13]</sup>;Wang 等人<sup>[14]</sup>提出一种社会化推荐算法,该算法综合考虑用户的信任关系、项目间的相似关系、评分信息,但是,其方法简单使用用户间的相似度替代信任度,整个计算过程只利用评分信息,信任关系未有被真正挖掘利用.

总之,面向大数据环境下的信任推荐,对信任信息的利用欠佳,预测准确性有待提升;且相似关系与信任关系的融合,需要高效统一的模型.因此,提出一种新的基于社交网络的推荐算法 Trust-PMF.算法在概率矩阵分解框架上进行扩展,并融入社交网络中的信任关系信息.主要贡献如下:

- 将现实中社交网络的信任关系引入推荐系统中,结合图综合考虑全局信任和局部信任,利用信任的传播特性对信任关系进行建模,以充分挖掘信任关系的结构性信息;
- 设置推荐权重,将主观信任和客观评分进行融合,综合考虑相似度和信任度来衡量用户的相似程度,由此提高识别邻居的能力,避免因数据稀疏导致邻居不相似问题,提高推荐的可信度和准确度,使得推荐结果更加契合用户的需求;
- 协同过滤思想,将社交网络的信任关系和评分信息融入概率矩阵分解模型,使用户对项目评分不仅受到自己的影响,同时也受到邻居用户的影响.并使用自适应权重动态决定各部分的影响程度,形成高效统一的模型;
- 实验在 FilmTrust 和 Epinions 这两个数据集上进行,我们的 Trust-PMF 算法与相关算法做了对比验证,同时评估和分析了算法中的参数对结果影响,实验显示算法的高效性.

本文第 1 节将简述信任的相关概念,介绍社会网络环境下基于信任的推荐算法的相关研究.第 2 节描述 Trust-PMF 算法原理、详细的推导过程,并给出 Trust-PMF 算法描述与分析.第 3 节进行实验及结果分析.最后,对研究工作进行总结和展望.

## 1 社交网络中的信任

用户之间因信任而形成社交圈,又通过社交来不断更新和强化信任.信任一旦建立,用户间的社交关系也逐

渐传递繁衍.因此,信任在社交网络中至关重要,它是社交形成的基础.社交网络中基于信任的推荐算法研究,关键就在于信任信息的挖掘和利用,因此首先介绍信任的相关概念.

### 1.1 信任的定义和性质

信任是复杂的概念,它随研究范畴的不同而有着不同的理解和定义.在计算机科学中,信任表示基于身份证、口令、密码系统等客体的契约和承诺.从推荐系统领域看,Golbeck 将信任定义为:假设用户  $B$  的行为会给用户  $A$  带来有利的参考和更好的结果,那么可以认定  $A$  信任  $B$ .图 1 是信任网络示例,1~6 这 6 个节点表示用户,节点之间的边为信任关系,边上的权重代表信任度的取值,如用户 1 指向用户 4 的边上有权重 0.4,表明用户 1 对用户 4 的信任程度为 0.4.

信任是一种复杂的网络关系,具有差异性、模糊性、传递性、非对称性、组合性、动态性等特点,具体描述如下.

- 差异性:信任是一种主观判断,受客观环境的影响,但由主体自身的情况决定,不同用户对于信任的评判标准是不统一的;
- 模糊性:信任度的大小可以用数值来表示和计算,但是信任本身具有不确定性,即,无法界定信任或者不信任的边值;
- 传递性:信任是可以传递的,如图 2 所示,假设用户  $A$  对  $B$ 、用户  $B$  对  $C$  有信任,那我们可以判定  $A$  在一定程度上也信任  $C$ ;
- 非对称性:信任关系是有方向性的,用户  $A$  信任  $B$ ,并不意味着反之亦然;
- 组合性:不相邻的两个用户,可以通过中间的传递路径来计算间接信任度.具体的组合方式可以根据算法不同而有所变化;
- 动态性:信任一旦建立之后并不是固定不变的,而是随时可能变化和更新.影响因素包括时间推移、领域知识的习得、客观环境的变迁、能力的提升等.

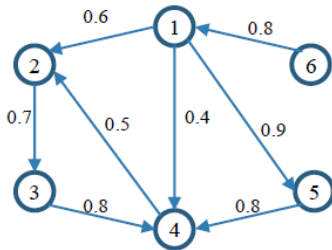


Fig.1 Trust network  
图 1 信任网络示意图

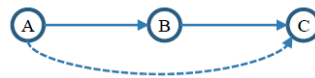


Fig.2 Transitive properties of trust  
图 2 信任的传递性

### 1.2 信任的分类和度量

信任的分类不同,度量方法也有所变化.可以从 3 个不同纬度对信任度进行分类,包括成对与成组方式、集中与分布方式、全局与局部方式.其中,全局和局部方式是目前最流行的分类方法,即,每个用户节点在整个网络中表现的全局信任度和对好友的局部信任度.

- 全局信任度:也叫做信誉,表示用户在社会网络中的信任参考了所有其他节点对其的综合评价<sup>[7]</sup>.一般用统计法和迭代法这两种方法来计算全局信任度:统计法记录信任或者不信任的数量,用比例等简单的公式表示;迭代法如 PageRank, EigenTrust 等;
- 局部信任度:表示两个用户之间一对一的信任,与其他用户对目标用户的看法以及态度无关,根据是否建立直接联系,可以分为直接信任和间接信任.

相比全局信任,局部信任的研究思路和方法多种多样,相对而言也更加复杂.两种局部信任度量方法如下.

- 基于信任的加权(TidalTrust)

用户  $a$  对某一物品的喜好程度,其值就是所有对该用户熟悉的用户  $u$  对物品  $i$  评分的均值.将这种方式作为基准的推荐策略叫做基于信任的加权平均,计算公式为

$$p_{a,i} = \frac{\sum_{u \in R^T} t_{a,u} r_{u,i}}{\sum_{u \in R^T} t_{a,u}} \quad (1)$$

其中,  $p_{a,i}$  表示用户  $a$  对项目  $i$  的预测评分,  $R^T$  表示评价物品  $i$  的邻居集合.  $t_{a,u}$  为用户  $a$  对  $u$  的信任,在这里也表示邻居用户对  $a$  的影响因子.最后的结果为了归一化需要除以用户  $a$  对所有邻居的信任度之和.

Golbeck 等人引进信任度  $t_{a,u}$  的估量方法,在基于信任的加权平均思想上改进得出 TidalTrust 推荐算法.

$$t_{a,u} = \frac{\sum_{v \in WOT_{(a)}^+} t_{a,v} t_{v,u}}{\sum_{v \in WOT_{(a)}^+} t_{a,v}} \quad (2)$$

式(2)就是公式(1)中信任值  $t_{a,u}$  的计量方式,通过宽度优先搜索所有能从起始节点到达目标节点的路径,并筛选出与目标节点之间存在最大信任度的用户集合,采取加权平均的方式综合考虑集合中所有用户对目标节点的信任,不断迭代更新.

- 基于信任的协同过滤(MoleTrust)

在协同过滤算法中,用户  $a$  对物品  $i$  的评分假设未知,则可以参考用户  $a$  的邻居用户的历史行为,通过对物品  $i$  的显式评分得到,如下式:

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in R^T} w_{a,u} (r_{u,i} - \bar{r}_u)}{\sum_{u \in R^T} w_{a,u}}$$

$\bar{r}_a$  表示用户  $a$  自己对所有物品评分的均值,预测评分除了此部分外,还考虑了邻居用户对它的影响.这里的邻居根据评分的相似程度得到.

用公式(2)计算的  $t_{a,u}$  替代权重  $w_{a,u}$ ,即将相似用户替换成信任用户,得到基于信任的协同过滤公式.

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in R^T} t_{a,u} (r_{u,i} - \bar{r}_u)}{\sum_{u \in R^T} t_{a,u}}$$

MoleTrust 算法是基于 TidalTrust 改进的,两者有很多共同点,例如都利用了信任传播属性,信任度的计算都采用加权平均的方式.但两者也存在差异,MoleTrust 的最大路径阈值是提前设定的,整个搜索过程是静态的;而 TidalTrust 的最大传播距离随着不同用户可以做出相应的调整和适当的变化.另外,在预测评分时,MoleTrust 使用协同过滤的方式,而 TidalTrust 仍然采用加权平均方法.

社会网络环境下基于信任的推荐算法 Trust-PMF:基于信任的概率矩阵因子分解模型.步骤如下.

- 首先,结合全局信任和局部信任,并利用信任的传递性质对信任关系进行建模;
- 然后设置推荐权重,将信任作为用户决策时的一个主观概念,结合客观存在的评分,综合考虑相似度和信任度来构建用户间的偏好关系;
- 接着,根据基于记忆的协同过滤思想将偏好关系融入概率矩阵因子分解(PMF)框架中,利用用户邻居交互的影响来弥补稀疏用户因训练数据不足导致的预测分数不准确问题;
- 最后,用梯度下降法更新未知的用户和项目的特征向量.

### 1.3 信任度及相似度计算

全局信任度在日常生活中更多地被称为声望、声誉、影响力等,表示用户在整个社交网络中的可靠性,由用户自身的诚信、能力等素质决定.全局信任度的计算公式(3)如下:

$$T_{global} = \frac{\ln d(t) - \text{Min}(\ln d(G))}{\text{Max}(\ln d(G)) - \text{Min}(\ln d(G))} \quad (3)$$

其中,

- $T_{global}$  为全局信任度;

- $\text{Max}(\text{In}d(G))$ 为图  $G$  中的最大信任入度, $\text{Min}(\text{In}d(G))$ 为最小信任入度,这两个值控制图连通的范围,是整个图性质的表征.用最大信任入度和最小信任入度的差值做规范化处理,可以使最后的取值控制在区间 $[0,1]$ 内;
- $\text{In}d(t)$ 表示目标用户  $t$  的信任入度,入度越大,即受到越多人的信任时,该用户在社会网络中的全局信任度越大,与真实场景吻合.

为了区分不同路径长度的信任,在 MoleTrust 算法基础上做出改进,得到局部信任的计算公式.

$$T_{local} = \frac{1}{d} t_{it} \tag{4}$$

$T_{local}$  是用户  $i$  对用户  $t$  的间接信任度,是根据 MoleTrust 算法得出的信任值. $d$  是根据宽度优先搜索算法得出的用户  $i$  和用户  $t$  的最短路径,当用户  $i$  到用户  $t$  的传播路径较长时,用户  $i$  对用户  $t$  表现出的局部信任越小.这一点在现实生活中也可以得到验证,即:当两个用户越亲密时,他们之间的信任关系也越强烈.

用户对其他用户的信任值是有权重的,根据被信任者的全局信任度和信任者的主观信任加以区分,综合考虑两者的影响,同时引入信任的传播特性来计算用户  $i$  对用户  $t$  的信任度  $T_{it}$ ,如公式(5)所示.

$$T_{it} = (1 - \beta) T_{global} + \beta T_{local} \tag{5}$$

其中, $\beta$ 的取值范围为 0 到 1,表示局部信任度所占的比重, $\beta$ 的值越大,说明在这个系统中,用户自身的主观信任逐渐成为两个用户信任程度的主导因素,而受信任者自身的影响力因素减弱. $\beta$ 最终的取值可以由系统的历史数据调试得出.

相似度的计算方法有很多,最简单的是欧几里德距离,其他常见的方法有相关相似性(皮尔逊相关系数)、余弦相似性和修正的余弦相似性等.

采用修正的余弦相似度进行计算:

$$\text{sim}(i, j) = \frac{\sum_{c \in I_{i,j}} (R_{i,c} - \bar{R}_i)(R_{j,c} - \bar{R}_j)}{(\sum_{c \in I_i} (R_{i,c} - \bar{R}_i)^2)^{1/2} (\sum_{c \in I_j} (R_{j,c} - \bar{R}_j)^2)^{1/2}} \tag{6}$$

其中, $I_i$  和  $I_j$  分别表示  $i$  和  $j$  进行评分过的推荐对象集合.

### 1.4 信任关系和相似关系的融合

通过设置推荐权重  $w_{it}$ ,自适应地平衡信任关系和相似关系的影响,可以提高识别邻居的能力,避免因数据稀疏导致邻居不相似问题,提高推荐的可信度.

$w_{it}$  即推荐权重,由下面的分段函数计算可得:

$$w_{it} = \begin{cases} 0, & n < n_1 \text{ 且 } T_{it} = 0 \\ T_{it}, & n < n_1 \text{ 且 } T_{it} \neq 0 \\ \frac{n}{n_2} S_{it}, & n_1 \leq n < n_2 \text{ 且 } T_{it} = 0 \\ \frac{n}{n_2} S_{it} + \left(1 - \frac{n}{n_2}\right) T_{it}, & n_1 \leq n < n_2 \text{ 且 } T_{it} \neq 0 \\ S_{it}, & n \geq n_2 \end{cases} \tag{7}$$

当用户  $i$  和用户  $t$  的共同评分  $n$  比系统最少评分值  $n_1$  还小时,可以忽略评分信息产生的影响;当共同评分数量  $n$  比系统最多评分值  $n_2$  还大时,说明两个用户选择的物品具有很高的重合度,推荐权重主要由评分相似性决定.当两个用户的共同评分值分布在区间 $[n_1, n_2]$ 内,则综合考虑信任关系和相似关系的影响,且相似度的影响权重由共同评分决定,影响因子为  $\frac{n}{n_2}$ .算法根据数据集的特征自适应地匹配计算方式,最后将筛选出  $w_{it}$  值最高的用户作为邻居集合.

### 1.5 可信推荐模型Trust-PMF

可信推荐模型 Trust-PMF 的建模,经过筛选出邻居用户后,将社交网络评分的信任信息和评分信息融入概率矩阵分解模型,使用户对项目评分不仅依赖于用户和项目潜在因子向量的交互,也受邻居集合评分的影响.并利用自适应权重 $\alpha$ 来动态决定各因素的影响程度,降低稀疏用户带来的噪声干扰.

Memory based 协同过滤思想是指当用户有相近的口味和偏好时,会形成相近的需求、表现类似的行为.用户有自己的口味,同时受到邻居用户的影响.即,用户  $u_i$  对物品  $v_j$  的预测评分为

$$R_{ij} = \alpha p_i q_j + (1 - \alpha) \frac{\sum_{t \in N_i} w_{it} p_t q_j}{\sum_{t \in N_i} |w_{it}|} \quad (8)$$

其中,  $w_{it}$  是公式(7)得到的邻居用户的推荐权值.用户对物品的评分  $R$  关于特征向量  $p, q$  的条件概率分布为

$$p(R | p, q, \sigma_R^2) = \prod_{i=1}^n \prod_{j=1}^m \left[ N(r_{ij} | \alpha p_i q_j + (1 - \alpha) \frac{\sum_{t \in N_i} w_{it} p_t q_j}{\sum_{t \in N_i} |w_{it}|}, \sigma_R^2) \right]^{I_{ij}^R}$$

其中,特征向量  $p, q$  也各自服从期望为 0、标准方差分别为  $\sigma_p^2 I, \sigma_q^2 I$  的球形高斯分布.

$$p(p | \sigma_p^2) = \prod_{i=1}^n N(p_i | 0, \sigma_p^2 I), p(q | \sigma_q^2) = \prod_{j=1}^m N(q_j | 0, \sigma_q^2 I).$$

扩展 PMF 模型,利用贝叶斯法则推导得到特征向量  $p, q$  的后验概率公式,并进行取  $\log$  处理.

$$\begin{aligned} \ln p(p, q | R, w, \sigma_R^2, \sigma_p^2, \sigma_q^2) = & -\frac{1}{2\sigma_R^2} \sum_{i=1}^n \sum_{j=1}^m I_{ij}^R \left( R_{ij} - g \left( \alpha p_i q_j + (1 - \alpha) \frac{\sum_{t \in N_i} w_{it} p_t q_j}{\sum_{t \in N_i} |w_{it}|} \right) \right)^2 - \frac{1}{2\sigma_p^2} \sum_{i=1}^n p_i^T p_i - \\ & \frac{1}{2\sigma_q^2} \sum_{j=1}^m q_j^T q_j - \frac{1}{2} \left( \sum_{i=1}^n \sum_{j=1}^m I_{ij}^R \right) \ln \sigma_R^2 - \frac{1}{2} (nl \ln \sigma_p^2 + ml \ln \sigma_q^2) + Con \end{aligned} \quad (9)$$

求公式(9)中对数函数的最大值,等价于求公式(10)损失函数的最小值.

$$L(R, w, p, q) = \frac{\lambda_p}{2} \|p\|_F^2 + \frac{\lambda_q}{2} \|q\|_F^2 + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^m I_{ij}^R \left( R_{ij} - g \left( \alpha p_i q_j + (1 - \alpha) \frac{\sum_{t \in N_i} w_{it} p_t q_j}{\sum_{t \in N_i} |w_{it}|} \right) \right)^2 \quad (10)$$

其中,  $\lambda_p = \sigma_R^2 / \sigma_p^2, \lambda_q = \sigma_R^2 / \sigma_q^2$ .

Trust-PMF 算法的设计:首先,基于图利用信任的传递性质对信任关系进行建模;然后,结合客观评分设置推荐权重,综合相似度和信任度的计算构建用户间的偏好关系,应用协同过滤将偏好关系融入 PMF 框架,利用邻居交互弥补数据稀疏导致的预测分数不准确;最后,用梯度下降法更新未知的用户和项目的特征向量.Trust-PMF 算法的伪代码见下面的算法.

**算法.** Trust-PMF.

Input: initial trust matrix  $T$ , rating matrix  $R$ , number of interactions  $n$ ,

$\alpha, \beta, n_1, n_2, learn\_rate, numNeighbors, numUsers m$ ;

Output: MAE, RMSE.

1. **for**  $i=1$  to  $m$  **do** //first step:calculate the recommending weight
2.     **for**  $j=1$  to  $m$  **do**
3.          $\{T_1 \leftarrow trust[i][j]\}$  //calculate local trust by formula (2)
4.     **for**  $i=1$  to  $m$  **do**
5.          $\{t_1 = \text{formula (3)}\}$  //calculate global trust
6.         **for**  $j=1$  to  $m$  **do**
7.              $\{T = (1 - \beta)t_1 + \beta t_1\}$  //calculate trust
8.              $trust[i][j] \leftarrow T$

```

9.     }}
10.    for  $i=1$  to  $m$  do
11.        for  $j=1$  to  $m$  do
12.            { $S \leftarrow sim[i][j]$  //calculate similarity by formula (6)}
13.            for  $i=1$  to  $m$  do
14.                for  $j=1$  to  $m$  do
15.                    { $w \leftarrow w[i][j]$  //choose final weight of recommend by formula (7)}
16.            initialize latent factor:  $p, q \sim N(x|0, \sigma^2)$  //second step:gradient descent
17.    for  $(i, j) \in TrainData$  do
18.        {  $p_i = p_i - learn\_rate \times \frac{\partial \mathcal{L}}{\partial p_i}$ 
19.           $q_j = q_j - learn\_rate \times \frac{\partial \mathcal{L}}{\partial q_j}$  }
20.     $predict(i, j)$  //third step:compute rating score by formula (8)
21.    Compute MAE and RMSE
22. }
```

算法分为 3 步实现:(1) 推荐权重的计算(第 1 行~第 15 行);(2) 梯度下降法求用户和项目的特征向量,计算推荐算法的性能衡量指标 MAE 和 RMSE(第 16 行~第 19 行);(3) 预测评分(第 22 行).算法分析中,核心的耗时在于公式的计算与迭代.

## 2 实验结果与分析

实验环境为 2.8GHZ Intel Core i5,8GB 内存,OS X Yosemite 操作系统.Trust-PMF 算法和对比算法均基于 JAVA 实现.采用了学术界广泛应用的平均绝对误差 MAE、均方根误差 RMSE 这两种指标来衡量推荐系统的性能.数据集选用真实数据集上进行实验,电影推荐网站 FilmTrust 和社会化电子商务网站 Epinions.数据集包含用户对项目的评分信息和用户之间的信任信息,数据集的基本特征信息见表 1.注意:为了统一评分区间,将两个数据集的评分通过公式  $\frac{r}{r_{max}}$  转化到[0,1].为计算方便,实验最大传播距离统一设定为 3.

Table 1 Statistics of data sets

表 1 数据集的统计特征

| DataSet         | Basic meta |         |                | User context |        |                |
|-----------------|------------|---------|----------------|--------------|--------|----------------|
|                 | Users      | Items   | Ratings (sale) | Density      | Users  | Links (type)   |
| FilmTrust (35K) | 1 508      | 2 071   | 35497[0.5,4.0] | 1.14%        | 1 642  | 1853 (trust)   |
| Epinions (385K) | 10 000     | 117 148 | 385122[1,5]    | 0.0349%      | 10 000 | 288146 (trust) |

### 2.1 实验结果及分析

为了验证 Trust-PMF 算法在推荐结果中表现出的优越性,从横向和纵向两个维度进行算法性能的评估.横向参考其他对比算法,纵向则对算法本身深入研究,分析和说明各个参数对推荐结果的影响.算法比较:

- PMF:Salakhutdinov 等人<sup>[15]</sup>提出的概率矩阵因子分解算法,算法仅利用了用户对项目的评分信息;
- MoleTrust:Massa 等人<sup>[4,5]</sup>提出的基于信任的协同过滤方法,该方法利用了信任的传递性;
- SocialMF:Jamali 和 Ester<sup>[11]</sup>提出的一种基于概率矩阵分解框架的经典算法,假设用户的偏好直接受到信任好友的影响,利用信任的传播性质计算目标用户的特征向量;
- SoRec:Ma 等人<sup>[9]</sup>首先提出的一种基于概率图的因子分析方法,该算法将信任网络结构和用户对项目的评分信息,通过共享的用户特征融入概率矩阵因子模型;

- RSTE:Ma 等人<sup>[10]</sup>提出的另一种基于社交网络的推荐方法,该算法认为,用户对项目的预测评分受到用户自身和信任用户两者的共同影响。

算法均随机选择 80% 的训练集进行训练,并行做 5 次交叉验证.其中,Trust-PMF 算法的参数设置为  $learn\_rate=0.01, \lambda_p=0.001, \lambda_q=0.01$ .表 2 分别表示数据集 FilmTrust 上,当特征向量的维度分别为 5 和 10 时,各种算法的准确度比较.经过大量的实验进行验证,对算法进行评估.

**Table 2 Accuracy comparison on FilmTrust**  
**表 2 FilmTrust 数据集不同算法的准确度比较**

| <i>dimension=5</i>  |         |          |         |         |           |           |
|---------------------|---------|----------|---------|---------|-----------|-----------|
| 准确度                 | PMF     | SocialMF | SoRec   | RSTE    | MoleTrust | Trust-PMF |
| MAE                 | 0.721 3 | 0.678 6  | 0.669 7 | 0.666 7 | 0.661 3   | 0.658 3   |
| RMSE                | 1.002 9 | 0.867 7  | 0.860 2 | 0.850 3 | 0.845 2   | 0.842 0   |
| <i>dimension=10</i> |         |          |         |         |           |           |
| 准确度                 | PMF     | SocialMF | SoRec   | RSTE    | MoleTrust | Trust-PMF |
| MAE                 | 0.77    | 0.670 8  | 0.673 9 | 0.668 8 | 0.656 8   | 0.655 6   |
| RMSE                | 1.061 8 | 0.853 5  | 0.858 9 | 0.858 4 | 0.870 1   | 0.848 9   |

从 FilmTrust 上的实验结果可看出:与仅利用评分信息而未融合信任信息的 PMF 算法相比,基于社会网络的推荐算法有较好的推荐准确性.说明信任信息的引入,能够弥补协同过滤算法因数据稀疏导致推荐准确性低的问题.相比于目前主流的几种基于社会网络的推荐算法,Trust-PMF 具有更好的推荐准确性.当维度为 5 和 10 时,其 MAE 以及 RMSE 值比其他算法都小.这是因为目前已有的算法一般仅仅停留在信任的表面信息而未充分挖掘朋友间深层的信任关系,Trust-PMF 权衡全局信任和局部信任的综合影响,同时考虑信任关系的传播特性,结合信任信息和评分信息,能够更好地对用户的行为和偏好进行建模.表 3 给出了 Epinions 上,当特征向量的维度分别为 5 和 10 时各个算法的准确度比较.

**Table 3 Accuracy comparison on Epinions**  
**表 3 Epinions 数据集不同算法的准确度比较**

| <i>dimension=5</i>  |         |          |         |         |           |           |
|---------------------|---------|----------|---------|---------|-----------|-----------|
| 准确度                 | PMF     | SocialMF | SoRec   | RSTE    | MoleTrust | Trust-PMF |
| MAE                 | 0.945 3 | 0.866 0  | 0.871 3 | 0.899 1 | 0.867 5   | 0.854 3   |
| RMSE                | 1.297 3 | 1.099 4  | 1.104 7 | 1.193 4 | 1.115 4   | 1.095 2   |
| <i>dimension=10</i> |         |          |         |         |           |           |
| 准确度                 | PMF     | SocialMF | SoRec   | RSTE    | MoleTrust | Trust-PMF |
| MAE                 | 1.055 6 | 0.872 9  | 0.877 3 | 0.912 6 | 0.887 5   | 0.866 5   |
| RMSE                | 1.303 0 | 1.141 1  | 1.165 3 | 1.274 3 | 1.145 4   | 1.137 7   |

上述在 FilmTrust 上的分析结果,同样适用于规模更大的数据集 Epinions.相比 PMF 算法和其他基于信任的推荐算法,Trust-PMF 算法在 MAE, RMSE 这两个指标上表现更好,推荐的准确度更高.说明信任信息的引入能够弥补协同过滤因数据稀疏导致推荐准确性低的问题.

## 2.2 参数的影响

本节分析迭代次数和预测评分对用户自身偏好的依赖程度  $\alpha$ 、信任度中局部信任的比重  $\beta$  以及两个用户的共同评分阈值 ( $n_1$  为下限阈值  $n_2$  为上限阈值) 对 Trust-PMF 算法推荐结果造成的影响.不妨设特征向量维度取 10.

由图 3、图 4 可知:Trust-PMF 算法随着迭代次数变化,MAE 和 RMSE 也随之发生改变.

刚开始,随着训练次数增加,推荐误差减少,之后逐渐趋于收敛.且在不同数据集上,达到最小误差时所需的迭代次数是不同的.当迭代次数为 80 和 40 时, FilmTrust 和 Epinions 上 Trust-PMF 的推荐准确度分别达到最高.

图 5、图 6 分别表示  $\alpha$  对 MAE 和 RMSE 的影响,其中,  $\alpha$  表示预测评分对用户自身偏好的依赖程度.  $\alpha$  的取值范围为 [0,1], 实验中依次取值为 0.0, 0.1, 0.2, ... 当  $\alpha=0.9$  时, 两个数据集上的推荐效果均达到最优. 此时, 最终的预测评分对好友偏好的依赖程度达到 10%, 对自身偏好的依赖程度为 90%. 另外, 当  $\alpha$  的取值从 0.7 增加到 0.8 时, 即, 最终的预测评分对邻居评分的依赖程度在 20%~30% 之间时, 两个数据集上预测的误差降低最快, 推荐效果提升



最为明显.

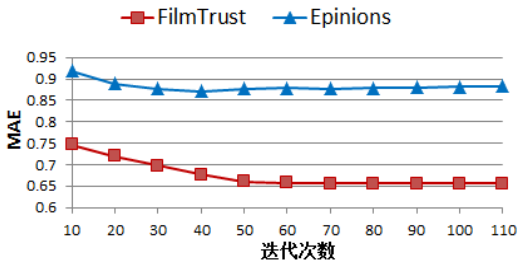


Fig.3 Effect of iteration on MAE  
图3 迭代次数对 MAE 的影响

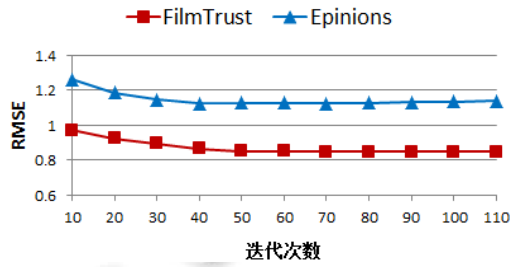


Fig.4 Effect of iteration on RMSE  
图4 迭代次数对 RMSE 的影响

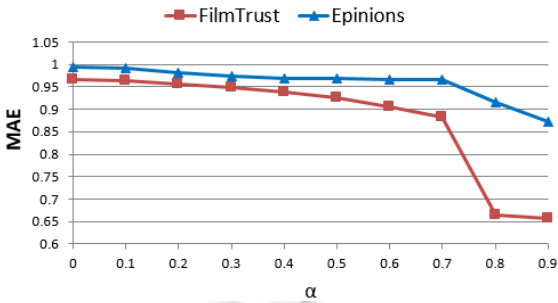


Fig.5 Effect of  $\alpha$  on MAE  
图5  $\alpha$ 对推荐 MAE 的影响

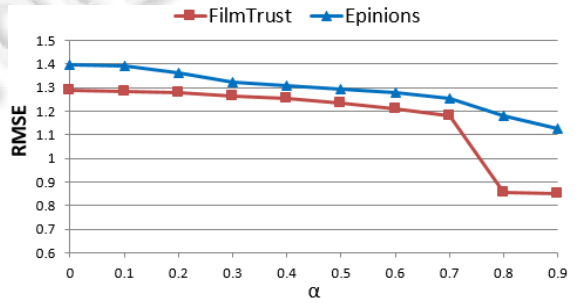


Fig.6 Effect of  $\alpha$  on RMSE  
图6  $\alpha$ 对 RMSE 的影响

图7和图8分别表示 $\beta$ 对MAE和RMSE的影响.Trust-PMF算法在不同数据集上达到最小预测误差时的 $\beta$ 取值是不同的,即,局部信任和全局信任在计算整体信任度时所占比重不同.当 $\beta$ 为0.1和0.3时,FilmTrust和Epinions上的推荐准确度分别达到最高,这可能是因为FilmTrust数据集上的信任关系信息要比Epinions数据集上的少很多.FilmTrust上一共只有609个用户真正存在直接信任关系,95%以上信任用户的好友数在10个以下,就是说,绝大部分用户的信任好友数量非常少.目标用户对其他用户很难形成主观的了解和认识,因此只能更大程度地依赖被信任好友自身的影响力来决定整体的信任程度.

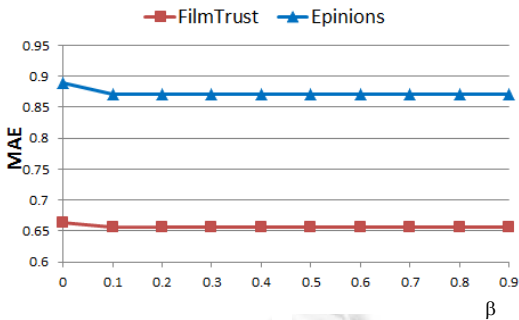


Fig.7 Effect of  $\beta$  on MAE  
图7  $\beta$ 对 MAE 的影响

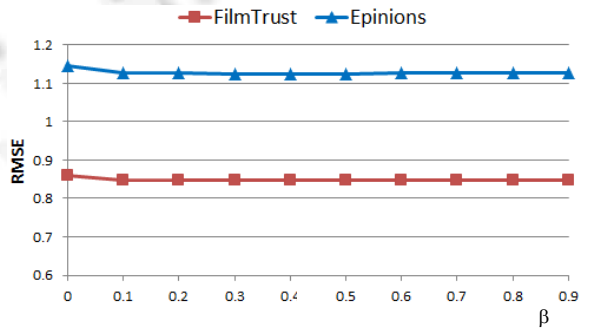


Fig.8 Effect of  $\beta$  on RMSE  
图8  $\beta$ 对 RMSE 的影响

图9和图10分别表示不同取值的 $n_2$ 对MAE和RMSE的影响. $n_1, n_2$ 分别表示推荐权重计算公式中的共同评分的上下阈值,为了确定这两个参数,假定 $n_1=0, n_2$ 从5开始取值,依次为5,6,7,8,...,15. $n_2$ 越大,Trust-PMF算法的预测误差越大.当 $n_2=5$ 时,Trust-PMF算法在两个数据集的准确度均达到最高.接着保持 $n_2=5$ ,使得 $n_1$ 从0

开始,取值为 0,1,2,3,4.当  $n_1$  为 0 时,两个数据集上推荐的误差达到最小.

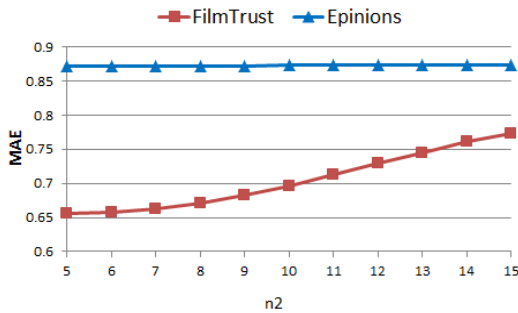


Fig.9 Effect of  $n_2$  on MAE

图 9  $n_2$  对 MAE 的影响

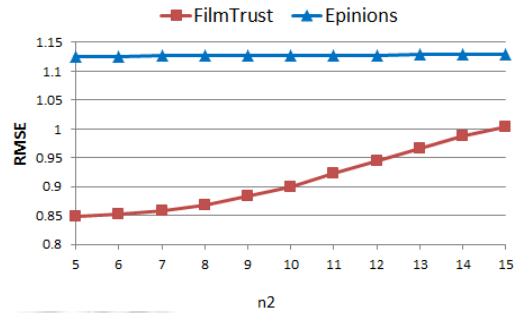


Fig.10 Effect of  $n_2$  on RMSE

图 10  $n_2$  对 RMSE 的影响

### 3 结束语

本文提出一种融合社交信息的推荐新方法 Trust-PMF.融合由评分数据产生的相似度和由信任关系信息产生的信任度构建用户的偏好模型生成邻居,再综合目标用户自身的偏好和邻居用户对其评分的影响预测评分,为基于信任的推荐研究提供了新思路.当然,社会网络中基于信任的推荐算法仍有继续研究和探索的空间,如考虑时间因素的信任关系动态模型、融合基于项目的协同过滤思想、融入社会网络中的上下文信息、探索不信任关系的推荐作用以及迁移到在线的大规模数据环境等.

### References:

- [1] Andersen R, Borgs C, Chayes J, Feige U. Trust-Based recommendation systems: An axiomatic approach. In: Proc. of the Int'l World Wide Web Conf. (WWW 2008). Beijing, 2008. [doi: 10.1145/1367497.1367525]
- [2] Kamvar SD, Schlosser MT, Garcia-Molina H. The eigentrust algorithm for reputation management in P2P networks. In: Proc. of the 12th Int'l Conf. on World Wide Web. ACM Press, 2003. 640–651. [doi: 10.1145/775152.775242]
- [3] Golbeck J. Computing and Applying Trust in Web-Based Social Networks [Ph.D. Thesis]. Maryland: University of Maryland, 2005.
- [4] Avesani P, Massa P, Tiella R. A trust-enhanced recommender system application: Mole skiing. In: Proc. of the 2005 ACM Symp. on Applied computing. 2005. 1589–1593. [doi: 10.1145/1066677.1067036]
- [5] Massa P, Avesani P. Trust-Aware recommender systems. In: Proc. of the 2007 ACM Conf. on Recommender Systems. Minneapolis, 2007. 17–24. [doi: 10.1145/1297231.1297235]
- [6] Jamali M, Ester M. Trustwalker: A random walk model for combining trust-based and item-based recommendation. In: Proc. of the 15th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Paris, 2009. 397–406. [doi: 10.1145/1557019.1557067]
- [7] Walter FE, Battiston S, Schweitzer F. A model of a trust-based recommendation system on a social network. Autonomous Agents and Multi-Agent Systems, 2008,16(1). [doi: 10.1007/s10458-007-9021-x]
- [8] Bedi P, Kaur H, Marwaha S. Trust based recommender system for the semantic Web. In: Proc. of the IJCAI 2007. 2007.
- [9] Ma H, Yang H, Lyu MR, King I. SoRec: Social recommendation using probabilistic matrix factorization. In: Proc. of the Int'l Conf. on Information and Knowledge Management. ACM Press, 2008. 931–940. [doi: 10.1145/1458082.1458205]
- [10] Ma H, King I, Lyu MR. Learning to recommend with social trust ensemble. In: Proc. of the 32nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. ACM Press, 2009. 203–210. [doi: 10.1145/1571941.1571978]
- [11] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proc. of the 4th ACM Conf. on Recommender Systems. 2010. 135–142. [doi: 10.1145/1864708.1864736]
- [12] Wang D, Ma J, Lian T, Guo L. Recommendation based on weighted social trusts and item relationships. In: Proc. of the 29th Annual ACM Symp. on Applied Computing. ACM Press, 2014. 254–259. [doi: 10.1145/2554850.2554884]

- [13] Guo L, Ma J, Chen ZM, Jiang HB. Incorporating item relations for social recommendation. Chinese Journal of Computers, 2014, 37(1):219–228 (in Chinese with English abstract).
- [14] Wang M, Ma J. A novel recommendation approach based on users' weighted trust relations and the rating similarities. In: Proc. of the Soft Computing. 2015. 1–10. [doi: 10.1007/s00500-015-1734-1]
- [15] Mnih A, Salakhutdinov R. Probabilistic matrix factorization. In: Proc. of the Advances in Neural Information Processing Systems. 2007. 1257–1264.

附中文参考文献:

- [13] 郭磊,马军,陈竹敏,姜浩然.一种结合推荐对象间关联关系的社会化推荐算法.计算机学报,2014,37(1):219–228.



陈婷(1991—),女,浙江金华人,硕士,主要研究领域为社会网络推荐服务.



周梦溪(1993—),男,硕士,主要研究领域为推荐算法,社交网络.



朱青(1963—),女,博士,副教授,博士生导师,CCF 高级会员,主要研究领域为大数据,图数据处理,分布式可信计算.



王珊(1944—),女,教授,博士生导师,CCF 会士,主要研究领域为高性能数据库,内存数据库,数据仓库.