

- $\text{Max}(\text{Ind}(G))$ 为图 G 中的最大信任入度, $\text{Min}(\text{Ind}(G))$ 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;
- $\text{Ind}(t)$ 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, β 的取值范围为 0 到 1,表示局部信任度所占的比重, β 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. β 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 α 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.        $\left\{ \begin{array}{l} p_i = p_i - learn\_rate \times \frac{\partial \mathcal{L}}{\partial p_i} \\ q_j = q_j - learn\_rate \times \frac{\partial \mathcal{L}}{\partial q_j} \end{array} \right\}$ 
19.      $predict(i, j)$  //third step:compute rating score by formula (8)
20.     Compute MAE and RMSE
21.   }

```

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^[15]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^[4,5]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^[11]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^[9]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^[10]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 α 、信任度中局部信任的比重 β 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 α 对 MAE 和 RMSE 的影响,其中, α 表示预测评分对用户自身偏好的依赖程度. α 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 α 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.

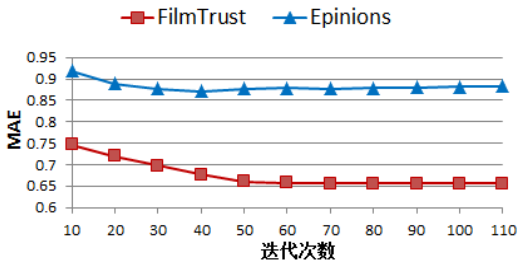


Fig.3 Effect of iteration on MAE
图3 迭代次数对 MAE 的影响

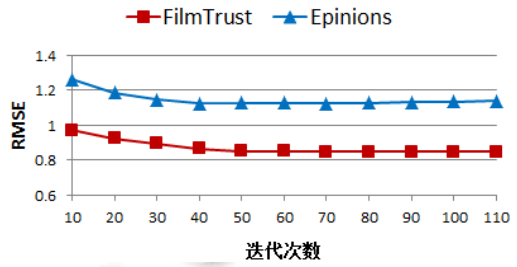


Fig.4 Effect of iteration on RMSE
图4 迭代次数对 RMSE 的影响

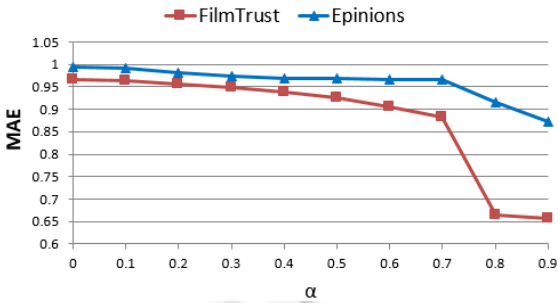


Fig.5 Effect of alpha on MAE
图5 alpha对推荐 MAE 的影响

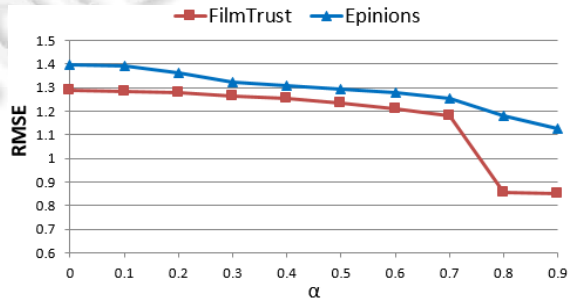


Fig.6 Effect of alpha on RMSE
图6 alpha对 RMSE 的影响

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.

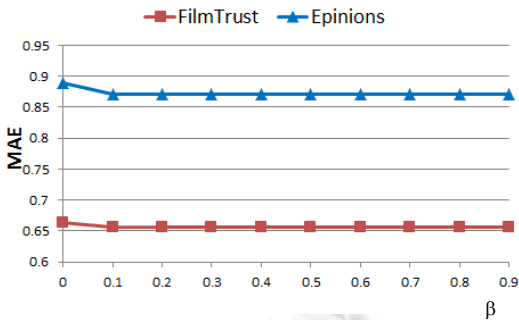


Fig.7 Effect of beta on MAE
图7 beta对 MAE 的影响

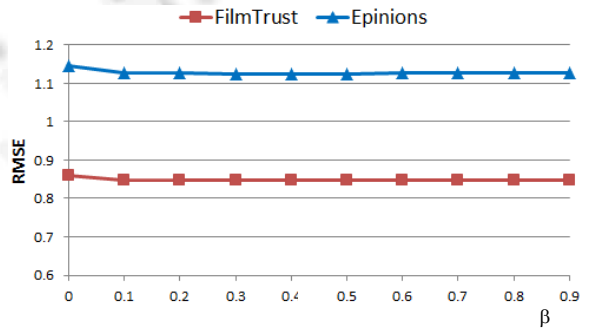


Fig.8 Effect of beta on RMSE
图8 beta对 RMSE 的影响

图9和图10分别表示不同取值的n2对MAE和RMSE的影响.n1,n2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n1=0,n2从5开始取值,依次为5,6,7,8,...,15.n2越大,Trust-PMF算法的预测误差越大.当n2=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n2=5,使得n1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.

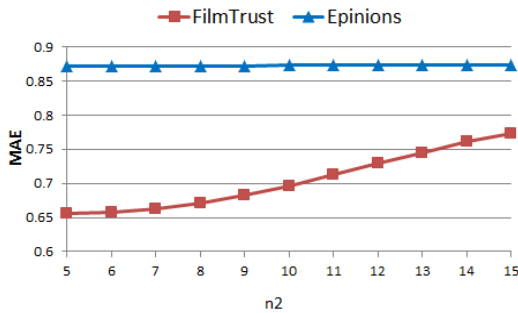


Fig.9 Effect of n_2 on MAE

图 9 n_2 对 MAE 的影响

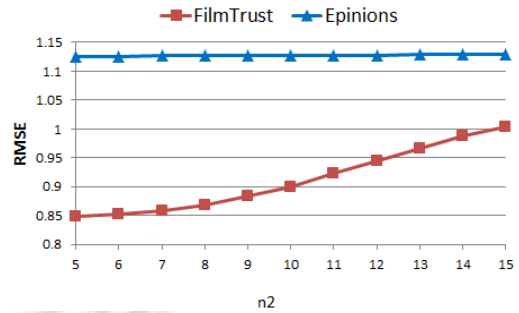


Fig.10 Effect of n_2 on RMSE

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]

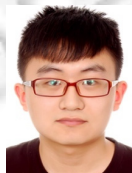
- [13] Guo L, Ma J, Chen ZM, Jiang HB. Incorporating item relations for social recommendation. Chinese Journal of Computers, 2014, 37(1):219–228 (in Chinese with English abstract).
- [14] Wang M, Ma J. A novel recommendation approach based on users' weighted trust relations and the rating similarities. 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.