

# 基于信任偏好的个性化推荐<sup>①</sup>

张秀杰

(天津理工大学 中环信息学院, 天津 300380)

**摘要:** 协同过滤推荐作为主流的个性化推荐方法在实际应用中存在一定缺陷, 在一些情况下得到的推荐结果不够准确. 考虑到信任与用户偏好相似性的关系, 将信任引入到推荐模型中, 并同时考虑暗示用户偏好的多维因素, 提出基于信任偏好的个性化推荐方法, 以提高推荐系统的准确性, 并用实验验证了此方法的有效性.

**关键词:** 信任; 偏好; 协同过滤; 个性化推荐; 推荐算法

## Personalized Recommendation Based on Trust and Preference

ZHANG Xiu-Jie

(Tianjin University of Technology, Zhonghuan Information College, Tianjin 300380, China)

**Abstract:** As the most popular personalized recommendation method, collaborative filtering recommendation has some shortcomings in real application, and the results of recommendation are not accurate enough in some cases. Taking into account the relationship between the trust and similarity of user preferences, we introduce trust into recommendation model, and take dimensional factor indicating user preferences into account, putting forward a recommendation method based on trust and preference. This method is aimed at solving the problem in collaborative filtering method and improving the accuracy of recommendation system. We have demonstrated the effectiveness of this method.

**Key words:** trust; preference; collaborative filtering; personalized recommendation; recommendation algorithm

### 1 引言

随着网络的飞速发展, 信息呈现出了爆炸式的增长. 人们如何在海量的信息中快速找到自己需要的资源, 成为研究的热点. 个性化推荐技术在此背景下应运而生. 现有的推荐技术大致分为以下几类: 协同过滤推荐, 基于内容的推荐, 基于人口统计的推荐, 基于效用的推荐, 基于知识的推荐, 基于规则的推荐. 其中最流行的是协同过滤推荐. 然而, 传统的协同过滤推荐技术存在一些问题, 比如冷启动、黑匣子、稀疏性等, 这些问题的存在使得推荐系统得到的结果在很多情况下不够准确.

为解决上述问题, 很多研究试图将信任模型引入到推荐系统中, 以替代或者补充现有的推荐技术. 如 Paolo Massa 等人<sup>[1]</sup>为解决传统协同过滤模型的数据稀疏性问题, 提出信任度量推荐系统; 张富国等人<sup>[2]</sup>考虑传统算法忽视推荐结果多样性这一问题, 建立的信

任推荐; Leonardo Zanette 等人<sup>[3]</sup>为解决存在于协同过滤推荐系统中的冷启动和评分问题, 将信任应用到推荐模型. 然而, 以上基于信任的推荐模型都存在不同程度的缺陷: Paolo Massa 等人的模型中, 用户通过对其他用户直接评分建立信任关系, 可操作性不强, 且容易造成用户的不满; 张富国等人的研究中对协同过滤系统中的稀疏性等问题则没有考虑; Leonardo Zanette 等人的研究, 用户偏好直接来自于用户打分这一单一数据, 缺乏准确性.

为解决存在于上述模型中的不足, 本文提出基于信任偏好的个性化推荐模型. 本研究是在 web2.0 环境下, 针对存在标签的系统提出来的. 已有研究表明<sup>[4]</sup>, 将标签应用在个性化推荐过程中, 能提高推荐结果准确性. 此模型在解决传统协同过滤存在的冷启动、稀疏性等问题的同时, 也将推荐结果的准确性进行了考虑. 在推荐模型中引入信任, 在一定程度上解决了冷

<sup>①</sup> 收稿时间:2013-05-24;收到修改稿时间:2013-07-05

启动、稀疏性等问題; 在用户兴趣度量上, 考虑了多维因素, 因此得到的兴趣列表将更加准确.

### 2 基于信任偏好的个性化项目推荐模型

基于信任偏好的个性化推荐算法与传统的协同过滤推荐方法的主要区别在于邻居用户的产生机理. 传统方法通过比较用户相似性的程度来确定邻居用户, 而基于信任的推荐方法通过比较用户间的信任度来确定邻居用户集. 信任模型的推荐步骤和 CF 方法大致相同: 第一步, 产生用户兴趣得分和邻居用户(信任关系); 第二步, 进行推荐. 其中不同的是邻居用户的产生方法和用户偏好的度量方法. 基于信任偏好的个性化项目推荐过程模型如图 1 所示.

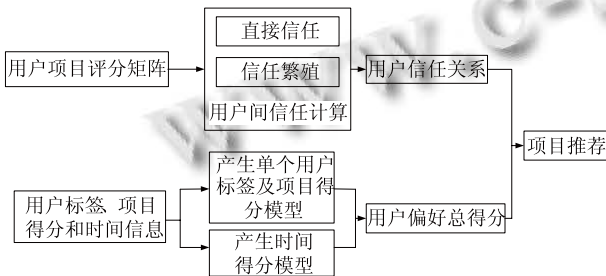


图 1 基于信任偏好个性化项目推荐过程模型

#### 2.1 直接信任用户的产生

此阶段建立在用户历史信息的基础上. 用户对已经进行过的评分的项目的评分值将被用于发现其信任用户的过程中, 信任用户将被作为当前用户的邻居用户.

两个用户之间的相关性由他们对相同项目的评分情况来计算, 本研究中, 用户与用户之间是通过标签联系起来的, 如图 2 所示. 所以, 本研究中用户对项目的评分由用户对此项目标注的标签评分的平均值来表示, 其结果可以表示为一个评分矩阵 A:

$$A(U \times I) = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,I} \\ r_{2,1} & r_{2,2} & \dots & r_{2,I} \\ \dots & \dots & \dots & \dots \\ r_{U,1} & r_{U,2} & \dots & r_{U,I} \end{bmatrix}$$

其中, U 代表用户, I 代表项目,  $r_{u,i}$  代表通过计算得到的用户 u 对项目 i 的评分, 即用户对项目标注标签评分的平均值.

SSS 使用上面评分矩阵中的数据, 即可按下面的式子计算用户之间的相似性:

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

其中, sim(u,v) 为用户 u 和用户 v 之间的相似性, m 为同时被用户 u 和用户 v 评过分的的项目,  $\bar{r}_u$  为用户 u 对其评过分的的项目的评分平均值,  $\bar{r}_v$  为用户 v 对其评过分的的项目的评分平均值.

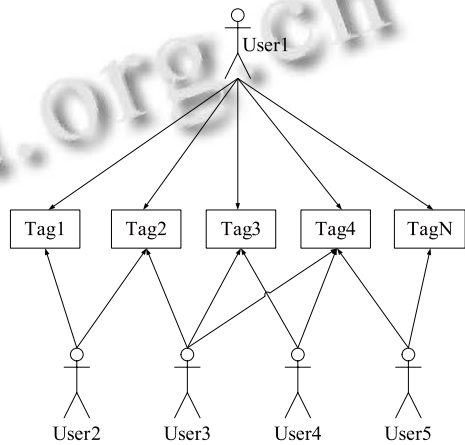


图 2 用户联系图

由此得出的用户相关性值是介于区间[-1,1]的, 并且只有同时被两个用户评分的项目被用于用户相关性计算的过程中.

用户之间相关性计算出来后, 需要采取一定的过滤方法选择出与当前用户 u 相关的 N 个用户作为其邻居用户, 即信任用户. 筛选策略可以是: 相关性排名靠前的前 N 个用户; 所有相关性值大于 0.3 的用户等等. 一般认为, 当用户 u 与用户 v 之间的信任度(trust)和用户 u 与用户 v 之间的相似度(similarity)同时可计算并具有足够的可信度(confidence)时, 两者的值是相似的<sup>[5]</sup>. 因此, 将用户 u 与用户 v 之间的信任度表示为  $T_{u,v}$ , 则可以定义  $T_{u,v} = sim(u,v)$ .

#### 2.2 信任繁殖

在实际环境中, 往往两个用户同时评价一个项目的情况并不普遍, 因而需要对存在的少数直接信任关系进行繁殖, 以此拓宽信任网络. 通过信任繁殖得到的邻居用户数量通常大于传统协同过滤推荐算法中得到的邻居用户的数量, 这样也正解决了传统的协同过滤推荐算法中邻居用户数量太少, 即冷启动的问题.

为拓宽信任关系,设计了一些规则.比如,如果用户  $A$  信任用户  $B$ , 用户  $B$  信任用户  $E$ , 则系统可以推断用户  $A$  信任用户  $E$ . 基于此, Golbeck<sup>[6]</sup> 创建了一个用户之间信任繁殖的算法. 用户  $A$  对用户  $E$  的信任可以表示为  $T_{A,E}$ , 如下式所示.

$$T_{A,E} = \frac{\sum_{u=1}^n \begin{cases} (T_{A,u} * T_{u,E}) & T_{A,u} \geq T_{u,E} \\ (T_{A,u})^2 & T_{A,u} < T_{u,E} \end{cases}}{\sum_{u=1}^n T_{A,u}} \quad (2)$$

其中,  $n$  表示用户  $A$  有  $n$  个直接信任的用户. 此模型计算中, 确保了信任繁殖过程中, 用户间的信任来源最重要的仍是用户之间的直接信任.

### 2.3 用户偏好的度量

考虑到作为反映用户偏好的重要因素, 用户的标签、用户的打分和用户标注的时间对提高用户偏好度量结果准确性均起到重要作用<sup>[4]</sup>, 本文在度量用户的偏好时, 借鉴综合考虑标签、得分和用户偏好时效性的个性化项目推荐模型. 这里计算的是当前用户的邻居用户的偏好情况.

依据考虑标签、得分和用户偏好时效性个性化推荐模型得到的用户偏好矩阵为  $M(A \times I)$ :

$$M(A \times I) = \begin{bmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,I} \\ M_{2,1} & M_{2,2} & \dots & M_{2,I} \\ \dots & \dots & \dots & \dots \\ M_{A,1} & M_{A,2} & \dots & M_{A,I} \end{bmatrix}$$

其中,  $M_{a,i}$  代表综合考虑了标签、项目得分和时间得分的用户  $a$  对项目  $i$  的喜欢程度, 其表达式为:

$$M_{a,i} = \lambda w(a,i) + (1-\lambda)w_{time}(a,i) \quad (3)$$

其中,  $w(a,i)$  代表考虑了标签和项目得分的用户  $a$  对项目  $i$  的评分情况;  $w_{time}(a,i)$  代表考虑了用户偏好可能会随时间的推移而变化这个因素得到的时间得分; 参数  $\lambda$  介于 0 和 1 之间, 如果更多的考虑用户的偏好, 则  $\lambda$  的取值更接近于 1, 如果更多的考虑用户偏好随时间的变化, 则  $\lambda$  的取值更接近于 0. 一般情况下, 可取  $\lambda=0.5$ , 即认为两者同样重要.

### 2.4 考虑直接信任用户的项目推荐

考虑用户对项目的偏好得分和当前用户  $A$  对其他用户的信任排名情况, 得到用户可能感兴趣的项目的推荐模型, 推荐得分  $score(A, i_x)$  表达式为:

$$score(A, i_x) = \frac{\sum_{v \in Neighbor(A)} M_{v,i_x} \times T_{A,v}}{|\sum_{v \in Neighbor(A)} T_{A,v}|} \quad (4)$$

其中,  $Neighbor(A)$  表示当前用户  $A$  的邻居用户集(直接信任用户);  $i_x$  为已经被用户  $v$  评分而未被当前用户  $A$  评分的项目;  $M_{v,i_x}$  为邻居用户  $v$  对项目  $i_x$  的喜欢程度;  $T_{A,v}$  为用户  $A$  对邻居用户  $v$  的直接信任度.

对于当前用户  $A$ , 项目推荐得分  $score(A, i_x)$  排名靠前的  $N$  个项目作为推荐项目.

### 2.5 考虑用户信任繁殖后的项目推荐

信任繁殖后, 用户的信任用户集发生了变化. 推荐得分  $score(A, i_x)$  可以表示为:

$$score(A, i_x) = \frac{\sum_{v \in Neighbor(A)} M_{v,i_x} \times T_{A,v}}{|\sum_{v \in Neighbor(A)} T_{A,v}|} \quad (5)$$

$Neighbor(A)$  为当前用户  $A$  经信任繁殖之后得到的信任用户集;  $T_{A,v}$  为信任繁殖之后, 当前用户  $A$  对用户  $v$  的信任度.

当推荐得分按照上述方法生成之后, 将项目推荐得分  $score(A, i_x)$  排名靠前的  $N$  个项目推荐给当前用户  $A$ , 这样就实现了对于用户  $A$  而言可能最感兴趣的项目的推荐.

## 3 基于信任偏好的推荐系统和传统协同过滤系统对比分析

在信息推荐领域, 协同过滤推荐是最成功的推荐技术<sup>[7]</sup>. 尽管如此, 基于相似性的协同过滤推荐技术还是存在一定的缺陷: 用户通常倾向于仅为少数的项目进行评分. 这样就造成用户之间的相似性很难计算. 另外, 传统的协同过滤推荐算法过多的强调了相似性, 而用户之间的其他特性, 比如信任, 没有得以体现. 然而实际上, 在现实世界中, 我们经常通过自己信任的用户获取项目(比如电影, 书籍等等)推荐的信息, 我们的兴趣很容易受自己信任的人所影响. 虽然已经有过通过用户声明信任用户的方法来协助项目推荐的过程的研究, 然而这些方法中, 用户间的信任没有被精确度量. 首先, 考虑到隐私问题, 用户不愿意在公共场合声明对他人的信任情况. 其次, 一些方法中仅仅考虑了用户之间的直接信任关系. 最后, 这些方法仅仅依赖用户的评分情况来建立信任关系, 进而进行项目的推荐, 而忽视了诸如标签、偏好时效性等与用户相关的其他的属性对

项目推荐准确度的影响. 基于信任偏好的推荐系统与传统协同过滤系统考虑因素对比如表 1.

表 1 基于信任偏好推荐系统与传统协同过滤推荐系统对比

系统类型 考虑因素	基于信任偏好推荐系统	传统协同过滤推荐系统
标签	考虑	不考虑
评分	考虑	考虑
偏好时效性	考虑	不考虑
信任	考虑	不考虑

引进信任的综合标签、得分和偏好时效性的个性化推荐系统, 旨在为协同过滤系统存在的问题提供可行的解决方案. 本文提出的基于信任的个性化项目推荐模型在邻居用户的产生上更加科学, 因为相比于含时效性的用户兴趣信息, 使用信任来度量用户之间的相似性更加准确; 同时此模型在单个用户的兴趣度量上同时考虑标签、得分和偏好时效性多维因素, 得到的偏好结果更加准确.

### 4 实验及分析

为验证基于信任偏好的个性化推荐算法的优越性, 设计如下实验, 将其与传统的协同过滤推荐算法进行比较.

#### 4.1 实验数据

实验数据来自于支持协同过滤标注系统的知名网站 CiteULike(<http://www.citeulike.org/faq/data.adp>)提供的数据集, 并根据本文的特殊需求, 使用 Matlab 模拟技术对数据不完整部分进行了模拟仿真, 包括 7221 个用户的 143870 个不同的标签和 434039 个不同的项目数据. 其中, 训练集占 80%, 测试集占 20%.

#### 4.2 度量标准

本文的最终目的是提高推荐结果的准确性, 所以主要考察算法的推荐准确度. 平均绝对偏差 MAE 通过计算用户对项目的预测评分和实际评分之间的偏差来度量算法的推荐准确性, MAE 越小, 推荐结果越准确.

将用户对项目的实际评分表示为集合  $\{r_1, r_2, \dots, r_n\}$ , 预测评分表示为集合  $\{s_1, s_2, \dots, s_n\}$ , 则 MAE 定义为:

$$MAE = \frac{\sum_{i=1}^N |s(i) - r(i)|}{N} \quad (6)$$

其中,  $N$  为推荐项目的个数.

对所有用户的 MAE 计算平均值, 得到总 MAE:

$$MAE = \frac{\sum_{i=1}^n MAE_i}{n} \quad (7)$$

### 4.3 结果分析

为验证本文推荐算法的有效性, 将其与传统协同过滤推荐方法进行对比.

表 2 为使用基于信任偏好的推荐方法得到的部分用户对部分项目的推荐总得分结果.

表 2 部分用户偏好总得分

I \ U	I1	I2	I3	I4	I5
U1	0.40		0.55	0.67	
U2		0.54		0.82	
U3	0.64		0.46	0.36	0.51
U4	0.57	0.48		0.67	
U5	0.5	0.61			
U6			0.75		0.5
U7	0.68		0.57		0.61
U8	0.75			0.5	
U9	0.61				0.5
U10	0.75	0.75			

传统协同过滤推荐方法中, 以夹角余弦相似性为度量方法. 用户邻居从 4 个增加到 32 个, 间隔为 4, 观察不同条件下的推荐效果, 实验结果如图 3 所示.

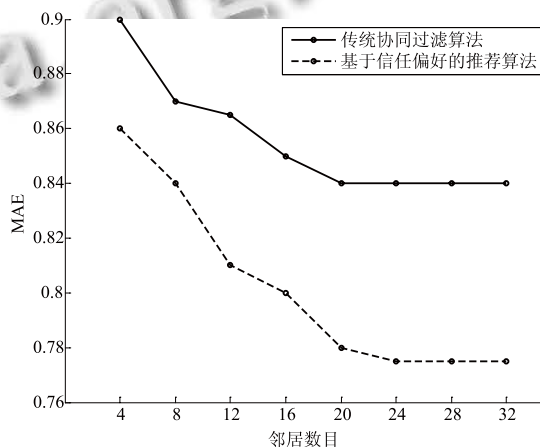


图 4 两种不同算法的性能比较

从实验结果可以看出, 当用户邻居较少时, 基于信任偏好的推荐方法比传统协同过滤模型有较好的推荐准确性, 说明该模型在存在数据稀疏性问题的系统

中也能取得较好的推荐结果. 随着邻居用户数量的增加, 基于信任偏好的推荐方法的优势越来越明显, 得到的结果更加准确. 从整体上来说, 基于信任偏好的个性化推荐方法具有较好的推荐准确性.

## 5 结论

本文论述了信任在个性化推荐系统中的应用情况以及现有系统的不足之处, 在此基础上提出基于信任偏好的个性化推荐算法, 解决传统协同过滤模型的冷启动、稀疏性和推荐准确性等问题, 以提高推荐结果的准确性. 最后, 用实验方法验证了此推荐方法的有效性.

## 参考文献

- 1 Massa P, Avesani P. Trust Metrics in Recommender Systems. *Human-Computer Interaction Series*, 2009: 259–285.
- 2 张富国, 徐升华. 基于信任的电子商务推荐多样性研究. *情报学报*, 2010, 29(2): 350–355.
- 3 Zanette L, Motta LRC, Santoro FM, Elia M. A Trust-based Recommender System for Collaborative Networks. *Proc. of the 2009 13th International Conference on Computer Supported Cooperative Work in Design*. 2009. 197–202.
- 4 张秀杰, 朱克珊, 李钢. 基于标签、得分和偏好时效性的项目推荐方法. *计算机系统应用*, 2012, 21(3): 202–205.
- 5 Massa P. Trust-aware Decentralized Recommender Systems [Ph.D. Thesis]. Italy: University of Trento, 2003.
- 6 Goldbeck J. Generating Predictive Movie Recommendations from Trust in Social Networks. *Int. Conf. on Trust Management*, Italy. 2006. 93–104.
- 7 Wang J, Yin J, Liu YZ, Huang CG. Trust-based Collaborative Filtering. *2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery*. 2011. 2650–2654.
- 8 Scaler JB, Konstan JA, John Riedl. E-Commerce Recommendation Applications. *Data Mining and Knowledge Discovery*. 2001. 115–153.
- frequency offset estimation for interleaved OFDMA uplink. *IEEE Trans. Commun.*, 2004, 52(9): 1585–1594.
- 5 Pun MO, Morelli M, Kuo CCJ. Iterative detection and frequency synchronization for OFDMA uplink transmissions. *IEEE Trans. Wireless Commun.*, 2007, 6(2): 629–639.
- 6 Cao ZR, et al. Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink. *IEEE Trans. on Communications*, 2004, 52(9): 1585–1594.
- 7 Pun M, et al. Maximum-likelihood synchronization and channel estimation for OFDMA uplink transmissions. *IEEE Trans. on Communications*, 2006, 54(4): 726–736.
- 8 Cao ZR, Tureli U, Yao YD. Frequency synchronization for generalized OFDMA uplink. *IEEE Global Telecommunication Conference, BLOBECOM 2004, 2007, 3: III-293-III-293*.
- 9 Cao Z, Tureli U, Yao Y. Deterministic multiuser carrier-frequency offset estimation for interleaved OFDMA uplink. *IEEE Trans. Commun.*, 2004, 52: 1585–1594.
- 10 Cao Z, Tureli U, Yao Y. Efficient structure-based carrier frequency offset estimation for interleaved OFDMA uplink. *Proc. 54th IEEE ICC Conf. Anchorage*. 2003. 3361–3365.
- 11 Sanguinetti L, Morelli M, Poor HV. Uplink Synchronization in OFDMA Spectrum-Sharing Systems. *IEEE Trans. on Signal Processing*, 2010, 58(5).
- 12 Proakis JG, Salehi M. *Digital Communications*. 5th ed., Beijing: Publishing House of Electronics Industry. 2012.

(上接第 135 页)