# 基于专家信任的协同过滤推荐算法

摘要

协同过滤算法通过已知评分估计未知评分，是推荐系统的重要组成部分。本文利用以信任为中心的社交网络，提出一种基于专家信任的协同过滤推荐算法, 该算法从普通用户中挖掘专家用户，得到专家因子，构建目标用户-专家用户相似度矩阵，利用近邻算法预测未知评分，改善了传统的协同过滤算法中由数据稀疏性和用户冷启动导致推荐准确率和覆盖率不高的问题。在FlimTrust数据集上的实验结果表明，该方法提高了推荐系统的准确率和覆盖率，有效缓解了用户冷启动问题。

**关键词**：推荐系统，信任，专家因子

**Collaborative filtering recommendation algorithm based on expert trust**

**Abstract**

Collaborative filtering algorithms estimate unknown scores by known scores and are an important part of the recommendation system. By analyzing the traditional collaborative filtering algorithm, the data accuracy and coverage are not high due to data sparsity and user cold start. This paper proposes a collaborative filtering recommendation algorithm based on trust by the trust-centered social network. The algorithm mines expert users from ordinary users, obtains expert factors, and builds a target user-expert user similarity matrix, which effectively reduces the sparseness of the data set and improves the accuracy of prediction. The experimental results on the FlimTrust dataset show that the proposed method improves the accuracy and coverage of the recommendation system and effectively alleviates the user's cold start problem.

**Keywords**: recommendation system ,trust,expert factor

## 1 引言

随着信息技术和互联网的迅速发展，人类逐渐进入信息过载的时代[1]，推荐系统应运而生。推荐系统根据用户的需求信息和行为特征为用户提供个性化服务，这在很大程度上解决了信息过载的问题。协同过滤算法通过对大量用户给出的评分协同处理进行推荐，是目前用于构建基于web的推荐系统的主流方法[2]。

根据推荐系统的组成部分，推荐系统可分为基于内容的推荐系统、协同过滤（CF）推荐系统和混合推荐系统[3]。协同过滤算法在推荐系统中应用最为广泛，它又分为基于用户的协同过滤和基于商品的协同过滤。本文中只研究基于用户的协同过滤，它的基本思想是根据相似用户（邻居）的兴趣爱好和历史行为，为目标用户推荐商品。因此，在协同过滤算法中识别相似用户（邻居）的能力是提高推荐质量最重要的部分[4,5]，而为了寻找相似用户，需要解决数据稀疏性和用户冷启动带来的计算相似度困难的问题。

为了解决以上问题，本文提出了一种基于专家信任的协同过滤推荐算法，在信任网络中挑选出此领域的专家用户，引入专家信任因子，计算目标用户-专家用户的相似度并完成推荐。在FlimTrust数据集上的测试结果表明该算法提高了推荐的准确率和覆盖率，在一定程度上缓解了数据稀疏性和用户冷启动问题。

本文第2节介绍了相关研究；第3节介绍了研究方法；第4节提出了基于专家信任的协同过滤推荐算法；第5节是实验结果与分析；最后是总结和展望。

## 2.相关研究

为了解决数据稀疏性和用户冷启动问题，国内外学者进行了大量的研究工作。文献[8]提出了一种新的计算相似度的方法，根据不同的目标用户选择不同的邻居。文献[9]将信任因素引入协同过滤推荐系统模型中，提高了推荐系统的准确度。文献[10]首次利用专家数据集计算相似度，提高了推荐质量。文献[11]提出了基于专家用户的可信推荐系统，利用信任网络挑选专家，在一定程度上解决了新用户的问题。

虽然，上述文献提到的方法在一定程度上缓解了数据稀疏性和用户冷启动问题，但还存在一些不足之处：1）考虑到有些领域的专家用户很匮乏，在数据稀缺的情况下无法直接利用专家数据集；2）在信任网络中得到专家用户的同时并没有考虑到将专家因子应用到相似度的计算中，对专家用户缺少进一步的挖掘；3）在信任网络中挑选专家用户时如何确定专家的阈值，这个范围尤其重要，不同的专家范围会影响推荐结果的准确性。考虑到以上问题，本文提出了一种基于专家信任的协同过滤推荐算法，首先基于信任网络挖掘出专家用户，然后调整阈值确定最佳专家用户，最后将专家因子引入到计算相似度中，从而缓解数据稀疏性和用户冷启动问题，提高预测的准确率和覆盖率。

## 3.研究方法

### 3.1 用户-商品评分矩阵

用户对商品的评分数据用m \* n 的矩阵表示，如表1所示，其中行表示有n位用户，列表示有m个商品，表示用户i对商品j的评分，１≤ｉ≤n，１≤ｊ≤m，的值在０～4.5分之间。若用户ｉ对商品ｊ没有评分，则＝０。

表1 用户-商品评分矩阵

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 用户 商品 |  | … |  | … |  |
|  |  | … |  | … |  |
| … | … | … | … | … | … |
|  |  | … |  | … |  |
| … | … | … | … | … | … |
|  |  | … |  | … |  |

### 3.2 相似度计算

在协同过滤算法中，计算用户之间的相似度主要有两种方法：余弦相似度和相关相似性。本文将分别采用这两种计算方法进行比较。

#### 3.2.1 余弦相似度

在基于余弦的方法[12]中，两个用户x和y被视为m维空间中的两个向量，可以通过计算它们之间角度的余弦值来测量两个矢量之间的相似性：

 （1）

其中代表用户x和y评价的所有商品集合。

#### 3.2.2 相关相似性

相关相似性又被称为Pearson相关相似性，用户x和y之间的Pearson相关系数[13]，[14]用公式(2)表示：

 （2）

其中，表示用户x和y共同评价商品的集合; 和分别表示用户x和y对商品评分的均值。

### 3.3基于用户的协同过滤算法

协同过滤算法是推荐系统领域中最早使用的方法之一。协同过滤算法在商业中取得了巨大的成就，其中包括著名的电子商务网站Amazon.com[16]。最早研究基于用户的协同过滤算法在文献[15]中，它的基本思想是：行为相似的用户往往会获得相似的评分。

基于用户的协同过滤算法的预测函数采用Resnick公式[17]，表示如下：

 （3）

式中表示用户u对商品j的预测评分；表示用户u对所有商品评分的均值；表示与目标用户u最相近的k位用户对商品j做出评分的用户集合；表示用户u和v之间的相似度。

### 3.4 基于专家的最近邻算法

文献[10]中提出了基于专家的最近邻算法，它是传统协同过滤算法的变形。它的基本思想是：使用专家评审用户，计算目标用户-专家用户的相似度，构建相似度矩阵，应用最近邻算法（KNN）选取与目标用户最近邻的k位专家，然后根据变形的Resnick公式预测目标用户的评分值。

计算用户u和专家e之间的相似度时，考虑了两个用户共同评定的商品数量，即调整因子，公式如下：

 （4）

其中，表示用户u和专家e共同评价商品的集合，表示用户u和专家e共同评定商品的数量，表示用户u评定商品的数量，表示专家e评定商品的数量。

应用KNN算法选取与目标用户u相似度最近邻的k位专家，组成专家集, 采用变形的Rsnick公式来计算目标用户对未知商品的预测评分：

 （5）

式中，表示目标用户u对所有商品评分的均值，表示专家e对商品评分的均值，表示用户u与专家e之间的相似度。

#### 4基于专家信任的协同过滤推荐算法

在可直接得到专家数据集的情况下，基于专家的最近邻算法通过直接计算目标用户与专家用户的相似度完成推荐。在现实生活中，很难直接得到专家用户集，需要从普通用户中挖掘专家用户。

### 4.1 专家用户评估模型

在实际生活中，每个领域都有专业知识较为全面的人，称这些人为专家用户。通常新用户更倾向于参考专家用户的建议，故他们的意见显得尤其重要。同样地，在推荐系统中，也存在一些“专家”用户。本文从用户评价商品的数量、拥有信任者的数量以及评分的客观性来定义推荐领域的“专家”用户。

#### 4.1.1活跃性

相较一般用户来说，专家用户专业知识更加全面，应该在评分过程中表现的更为积极，故专家用户评价商品的数量更多，计算用户活跃性的公式如下：

 （6）

其中，表示用户的活跃性，表示用户u评价商品的次数，表示评价商品最多的用户的评价次数。

#### 4.1.2 可靠性

信任度在信任网络的推荐中起到了关键作用[6,7]，专家用户理应拥有更多的信任人数，计算用户可靠性的公式如下：

 （6）

其中，表示用户的可靠性，表示用户拥有信任者的数量，表示拥有最多信任者的用户的信任者数量。

#### 4.1.3 客观性

专家用户在评价商品时应该更加客观和准确，评分应该更加接近商品的平均评分。假如商品i被N个用户评分，则商品i的平均评分为：

 （7）

将用户对每件商品评分偏差归一化处理后，使用用户评分偏差的平均值来表示用户的客观性，公式如下：

 （8）

式中F是用户u给出的评价次数；是商品i被评分的最大值。

#### 4.1.4 挑选专家

综合用户的活跃性、可靠性、客观性，定义专家用户的公式如下：

 （9）

其中代表各部分所占权重。

### 4.2 筛选最佳专家用户

寻找最佳专家用户的重要途径包括如何设定活跃性、可靠性、客观性的权重以及的阈值。

#### 4.2.1 设置影响专家用户因素的权重

根据定义专家的公式，的值会影响筛选出的专家用户，令，以0.1为间隔设置权重，确定使推荐准确率最高的权重值。

#### 4.2.2 设置调节专家数量的阈值

根据专家用户评估公式，当大于设置阈值，该专家用户为最佳专家用户e，存储最佳专家用户的专家值，构成最终的专家数据集。

### 4.3 引入专家因子的相似度

本文实现了从普通用户中挖掘出专家用户集，并将专家因子应用到推荐系统中，在文献[10]的基础上提出一种改进的计算用户相似度的方法，公式如下：

 （10）

其中，表示用户u与专家e之间的相似度，是比例调节因子，表示专家因子。

### 4.4 预测目标用户评分

建立目标用户-专家用户相似度矩阵，根据（10）所计算的相似度大小进行排序，由KNN算法得到目标用户u的最近邻专家用户集，则用户u对商品i的预测评分如式（11）所示：

 （11）

其中，表示用户u对所有商品评分的均值，表示最近邻专家e对商品评分的均值，表示用户u与专家e之间的相似度。

本文提出了基于专家信任的协同过滤推荐算法，该算法改进了传统的协同过滤算法，缓解了数据稀疏性和用户冷启动问题，具体步骤如下：

⑴建立专家用户评估模型，通过调整各因素权重和阈值确定最佳专家用户；

⑵根据改进的相似度公式，建立目标用户-专家用户的相似度矩阵；

⑶将相似度从小到大依次排序，找出最近邻的K位专家作为目标用户的邻居；

⑷根据预测评分公式，预测目标用户对未知商品的评分值。

## 5. 实验结果与分析

### 5.1．实验数据集

实验采用FlimTrust数据集[21]对算法的有效性进行实验。FlimTrust数据集来自一个基于社交网络的电影推荐系统。该数据集由两部分组成，用户评分集和信任集。

用户评分集中包括1508位用户对2071个商品的35494条评分数据，评分值为1~4.5中的任意值，评分值越大，表示用户越喜欢该商品。用户-商品评分数据集的稀疏度为98.8635%。

信任集中包括用户之间的1853条信任关系评分数据，评分值为1或者空，1表示用户的信任评分值，空值表示用户之间没有信任评分。

### 5.2. 评估标准

#### 5.2.1．精确率

精确率是评估推荐系统过程中最基本的指标之一。本文分别采用平均绝对误差MAE和均方根误差RMSE作为精确率度量标准[18]。

平均绝对误差MAE通过比较预测值与用户实际评分值之间的偏差来衡量预测的准确性。MAE越小，表明推荐结果越精确。假设预测的用户评分集合表示为，对应的实际用户评分集合是，n表示所有评分的数目，则计算平均绝对误差MAE的公式如下：

 （12）

均方根误差是预测值与真实值偏差的平方与观测次数n比值的平方根，均方根误差也是用来衡量预测值同真实值之间的偏差，偏差越小，表明推荐质量越高。设表示目标用户对第i件商品的预测评分值，表示目标用户对第i件商品的真实评分值，n表示所有评分的数量，则均方根误差RMSE的定义如下：

 （13）

### 5.2.2 覆盖率

在实际推荐过程中，即使一个推荐系统有较高的准确率，但仍然不能对其中某一部分商品或者用户做出推荐，这就需要引入覆盖率[19]。根据文献[20]，覆盖率主要分为两部分：1）预测覆盖率：推荐系统能够生成推荐的商品数量与所用商品数量的百分比；2）商品覆盖率：在生成推荐列表的情况下有效推荐给用户的可用商品的百分比。本文只对目标用户进行预测评分，故采用预测覆盖率。

预测覆盖率是指预测评分的数目与所用评分的数目之比，公式如下：

 （14）

### 5.3. 实验设计及结果分析

本实验的目的是将基于专家信任的协同过滤推荐算法与基于用户的协同过滤算法以及基于信任网络的推荐算法[22]进行比较分析。实验将Filmtrust数据集按照8:2的比例分为训练集和测试集。

（1）确定专家评估模型中的权重

实验将余弦相似度作为相似性度量，将权重分别设定为，计算平均绝对误差MAE的值，得到使平均绝对误差MAE值最小的权重值的结果：。

（2）专家值分布图

由公式（8）可以计算出Exp(u)值，分布如图1所示：

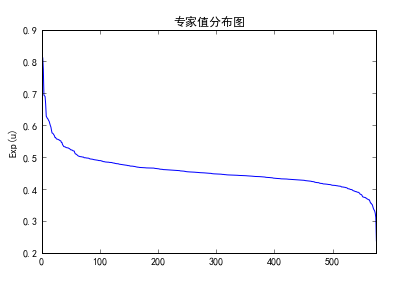


图1中的纵轴代表Exp(u)的值，可以看出Exp(u)值的分布范围在（0.2,0.9）。

（3）确定最佳专家用户阈值

使用余弦相似度作为相似性度量，列出所有专家用户阈值范围，得到如图2所示结果：

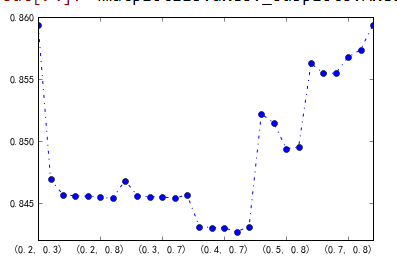


图2中的横坐标代表专家用户阈值范围，从左到右依次是（0.2,0.3），（0.2,0.4），（0.2,0.5），（0.2,0.6），（0.2,0.7），（0.2,0.8），（0.2,0.9），（0.3,0.4），（0.3,0.5），（0.3,0.6），（0.3,0.7），（0.3,0.8），（0.3,0.9），（0.4,0.5），（0.4,0.6），（0.4,0.7），（0.4,0.8），（0.4,0.9），（0.5,0.6），（0.5,0.7），（0.5,0.8），（0.5,0.9），（0.6,0.7），（0.6,0.8），（0.6,0.9），（0.7,0.8），（0.7,0.9），（0.8,0.9），纵坐标代表平均绝对误差MAE值。从结果看出，最佳专家用户阈值范围设定为（0.4,0.8），平均绝对误差MAE最低，推荐结果最好。

（4）确定合适的邻居k和比例调节因子

使用余弦相似度作为相似性度量，将分别设定为。当k分别为10,20,30,40,50时，可得到如图3所示预测结果：

（5）各种算法的比较

为了表明基于专家信任的协同过滤推荐算法的确提高了推荐系统的质量。我们将算法分为以下三组。第一组包括使用基于用户的协同过滤推荐算法，第二组包括使用中基于信任网络的推荐算法[22]，第三组包括使用基于专家信任的协同过滤推荐算法。

A.两种基于用户的协同过滤推荐算法，CF\_P和CF\_C：这些算法分别使用Pearson相关系数，余弦相似度来计算用户相似性，预测目标用户评分。

B.基于信任网络的推荐算法，CF\_T：该算法使用信任传递来预测目标用户评分。

C.两种基于专家信任的协同过滤推荐算法，CF\_E\_P，CF\_E\_C：这些算法分别使用Pearson相关系数，余弦相似度并且引入专家因子来计算相似性，预测目标用户评分。

实验结果如图4所示：

## 结束语

本文针对目前传统协同过滤算法存在的数据稀疏性和用户冷启动问题提出一种基于专家信任的协同过滤算法。该算法以信任为中心在社交网络中挖掘专家用户，并将专家因子融合到相似度计算中。在FlimTrust数据集上的实验结果验证了该算法的有效性。该算法相比较传统的协同过滤算法和基于信任网络的推荐算法，不仅提高了推荐系统的精确率和覆盖率，同时有效缓解了数据稀疏性和用户冷启动问题。

尽管本文提出的算法已经显示出它的有效性，但也为未来的改进提供了机会。首先，如何更加有效地从信任网络中挖掘专家；其次，该算法在不同数据集上如何选取最优参数；再者，专家用户也可以和各种CF结合使用，以提高推荐的准确性。

**参考文献**

[1 ] C. S. Saunders and A. Rutkowski, "Growing Pains with Information Overload," in Computer, vol. 43, no. , pp. 96, 94-95, 2010. doi:10.1109/MC.2010.171

[2] A. Tuzhilin and G. Adomavicius, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," in IEEE Transactions on Knowledge & Data Engineering, vol. 17, no. , pp. 734-749, 2005.

[3] [Marko Balabanović , Yoav Shoham, Fab: content-based, collaborative recommendation, Communications of the ACM, v.40 n.3, p.66-72, March 1997](https://dl.acm.org/citation.cfm?id=245124) [doi>[10.1145/245108.245124](http://doi.acm.org/10.1145/245108.245124)]

[4] J. Bobadilla, F. Ortega, A. Hernando, J. Alcala, Improving collaborative filtering recommender system results and performance using genetic algorithms,Knowledge-Based Systems 24 (2011) 1310–1316.  
[5] J. Bobadilla, F. Serradilla, J. Bernal, A new collaborative filtering metric that improves the behavior of recommender systems, Knowledge-Based Systems 23 (2010) 520–528.

[6]R. Levien. Attack-resistant trust metrics. Computing with Social Trust,Spring,pp.121-132,2009

[7]C.Ziegler ,S. McNee，J.Konstan,and G.Lausen.Improving recommendation lists through topic diversification. World Wide Web Conference,pp. 22-32,2005

[8] H.J. AhnA new similarity measure for collaborative filtering to alleviate the new user cold-starting problem Information Sciences, 178 (2008), pp. 37-51

[9] [John O'Donovan , Barry Smyth, Trust in recommender systems, Proceedings of the 10th international conference on Intelligent user interfaces, January 10-13, 2005, San Diego, California, USA](https://dl.acm.org/citation.cfm?id=1040870) [doi>[10.1145/1040830.1040870](http://doi.acm.org/10.1145/1040830.1040870)]

[10] [Xavier Amatriain , Neal Lathia , Josep M. Pujol , Haewoon Kwak , Nuria Oliver, The wisdom of the few: a collaborative filtering approach based on expert opinions from the web, Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, July 19-23, 2009, Boston, MA, USA](https://dl.acm.org/citation.cfm?id=1572033) [doi>[10.1145/1571941.1572033](http://doi.acm.org/10.1145/1571941.1572033)]

[11] 王宁,徐锋.基于专家用户的可信推荐系统[C].//2011全国软件与应用学术会议(NASAC2011)论文集.南京大学,2011:205-210.

[12] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-Based Collaborative Filtering Recommendation Algorithms,” Proc. 10th Int’l WWW Conf., 2001.

[13] U. Shardanand and P. Maes, “Social Information Filtering:Algorithms for Automating ‘Word of Mouth’,” Proc. Conf. Human Factors in Computing Systems, 1995.

[14] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl,“GroupLens: An Open Architecture for Collaborative Filtering of Netnews,” Proc. 1994 Computer Supported Cooperative Work Conf.,1994.

[15]J. Breese，D.Heckerman，and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. Conference on Uncertainty in Atificial Intelligence,1998

[16] G. Linden, B. Smith, and J. York, “Amazon.com Recommendations: Item-to-Item Collaborative Filtering,” IEEE Internet Computing, Jan./Feb. 2003.

[17]Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of ACM CSCW’94 Conference on Computer Supported Cooperative Work, Sharing Information and Creating Meaning, pages 175–186, 1994.

[18] T. Chai and R. R. Draxler. Root mean square error (RMSE) or mean absolute error (MAE)? –Arguments against avoiding RMSE in the literature. Geoscientific Model Development,7(3)，pp. 1247-1250，2004

[19] Herlocker J., Konstan J., Terveen L. and Riedl J. 2004. Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems, 22(1), pp. 5–53.

[20] [Mouzhi Ge , Carla Delgado-Battenfeld , Dietmar Jannach, Beyond accuracy: evaluating recommender systems by coverage and serendipity, Proceedings of the fourth ACM conference on Recommender systems, September 26-30, 2010, Barcelona, Spain](https://dl.acm.org/citation.cfm?id=1864761) [doi>[10.1145/1864708.1864761](http://doi.acm.org/10.1145/1864708.1864761)]

[21] <https://www.librec.net/datasets.html>

[22] Golbeck, J. and Hendler, J. 2006. FilmTrust: Movie recommendations using trust in Web-based social networks. In Proceedings of the IEEE Consumer Communications and Networking Conference. Las Vegas, NV.]]