

一种基于信任网络的协同过滤推荐策略^①

卢竹兵, 唐 雁

西南大学 计算机与信息科学学院, 重庆 400715

摘要: 提出了一种基于信任网络的协同过滤推荐策略, 在传统协同过滤策略中引入信任网络, 将相似度和信任度结合在一起, 提高推荐的准确率. 实验证明, 在数据稀疏的情况下该策略比传统的协同过滤推荐策略有更好的推荐效果.

关键词: 协同过滤; 推荐系统; 信任网络; 信任度

中图分类号: TP391

文献标识码: A

传统的协同过滤推荐系统存在 Sparsity 问题^[1], 即在数据稀疏情况下, 一方面, 若相似度无法计算, 系统就无法对这些用户进行推荐^[2]; 另一方面, 当用户间共同评价的项目数量很少时, 计算出的相似度也存在一定的误差, 影响推荐的准确率^[3,4]. 对此, 研究者提出了很多解决方案, 如对用户项目评价矩阵降维^[5]、运用关联信息获取相关联的项目^[6,7]、结合项目间相似度与用户的相似度^[8]等, 这些方法虽然都实现了更好的推荐, 但是到目前为止, 仍没有一种通用的克服 Sparsity 问题的方法. 文献[4]利用信任度的传递, 为目标用户匹配到更多的邻居, 可以在一定程度上解决协同过滤推荐存在的 Sparsity 问题, 而且有益于 Cold Start 用户, 但只给出了定性的分析, 没有给出具体的信任度传递规则. 本文根据实际情况, 定义了一组信任度传递规则, 并以此为基础为数据稀疏用户匹配到更多的邻居, 以提高系统的推荐准确率.

1 信任网络

信任网络是由系统中所有用户的信任关系所构成的网络^[2]. 如图 1 所示, 信任网络是一个有向网络, 网络中每个节点代表一个用户, 每条边表示用户间的信任关系, 边的权重为信任的程度.

由于系统中用户数量很多, 用户间有直接信任关系的数量很少, 表示信任网络的矩阵往往非常稀疏. 为了使原本不关联的用户通过信任关系相互联系, 可以通过信任关系的传递来实现. 如图 1 所示, 根据用户 A 与用户 B 以及用户 B 与用户 D 的直接信任关系, 可以推断用户 A 与用户 D 存在一定的间接信任关系.

信任度是指系统中用户之间的一种主观信任程度, 它可以分为两种, 一种是直接信任度, 另一种是间接信任度. 在信任网络中, 直接信任度是用户之间直接进行信任评价的值, 间接信任度是通过直接信任度的传递而建立起的信任程度.

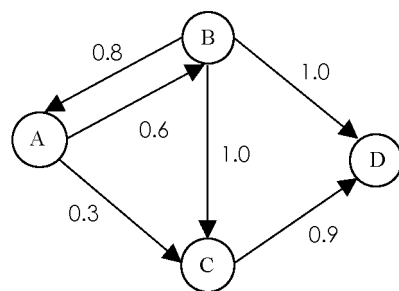


图 1 信任网络

① 收稿日期: 2007-11-07

作者简介: 卢竹兵(1981-), 男, 安徽安庆人, 硕士研究生, 主要从事 WEB 应用技术的研究.

通讯作者: 唐 雁, 教授, 硕士生导师, 主要从事智能科学研究.

2 基于信任网络的协同过滤推荐策略

针对传统协同过滤中所存在的 Sparsity 问题, 结合信任度在信任网络中的传递特性, 提出了一种基于信任网络的协同过滤推荐策略, 整体框架如图 2 所示.

在此策略模型中, 用户的评价矩阵和信任网络是原始输入数据, 工作流程如下: ①运用信任度的传递规则对原始的信任网络矩阵进行处理, 计算传递后的间接信任度, 保存到信任网络矩阵中, 得到相对稠密的信任度矩阵; ②运用用户评价矩阵计算用户间的相似度, 得到一个相似度矩阵; ③结合信任度矩阵和相似度矩阵, 产生 K 个最近邻居; ④生成推荐结果.

2.1 间接信任度的计算

采用信任度传递规则, 计算用户之间的间接信任度, 使原本稀疏的信任网络矩阵变得相对稠密.

计算规则如下:

① 若信任网络中节点 S 与节点 M 之间存在唯一路径 $C(S, N_1, N_2, \dots, N_i, M)$, 则: 节点 S 对节点 M 的间接信任度为在信任网络中节点 S 到节点 M 所有直接信任度的最小值.

$$T_{S \rightarrow M} = \text{MIN}(T_{S \rightarrow N_1}, T_{N_1 \rightarrow N_2}, \dots, T_{N_i \rightarrow M}) \quad (1)$$

② 若信任网络中节点 S 与节点 M 之间存在多条路径 $(C_1, C_2, C_3, \dots, C_i)$, 则节点 S 对节点 M 的间接信任度值为在信任网络中 S 到 M 经过每一条路径所得到的间接信任度的平均值.

$$T_{S \rightarrow M} = \text{AVG}(T_{S \rightarrow M}^{C_1}, T_{S \rightarrow M}^{C_2}, \dots, T_{S \rightarrow M}^{C_i}) \quad (2)$$

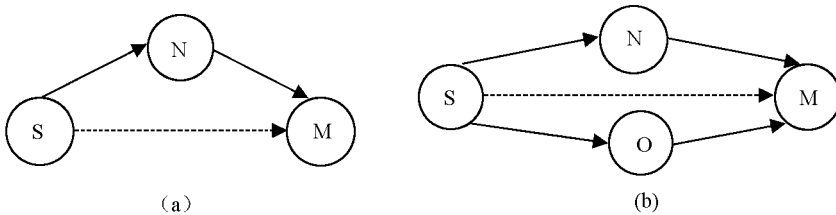


图 3 两种不同路径

例如: 如图 3(a)所示, 若节点 S 到节点 M 只有一条路径 $C(S, N, M)$, 则 S 对 M 的间接信任度 $T_{S \rightarrow M}$ 为 $T_{S \rightarrow M}^C$, 即 $T_{S \rightarrow N}$ 和 $T_{N \rightarrow M}$ 的最小值. 如图 3(b)所示, 若节点 S 到节点 M 有 2 条路径 $C_1(S, N, M)$ 和 $C_2(S, O, M)$, 则 S 到 M 的间接信任度 $T_{S \rightarrow M}$ 为 $T_{S \rightarrow M}^{C_1}$ 和 $T_{S \rightarrow M}^{C_2}$ 的平均值.

2.2 相似度的计算

相似度的计算方法有很多, 本文选用最常用的皮尔森相关性系数 (Pearson Correlation Coefficient)^[3,9], 公式如下:

$$\text{Similarity}(c, p) = \frac{\sum_{i=1}^m (c(i) - \bar{c}) * (p(i) - \bar{p})}{\sqrt{\sum_{i=1}^m (c(i) - \bar{c})^2 * \sum_{i=1}^m (p(i) - \bar{p})^2}} \quad (3)$$

其中, m 代表用户 c 与用户 p 共同评价的项目个数, $c(i)$ 表示用户 c 对项目 i 的评价值, $p(i)$ 表示用户 p 对项目 i 的评价值, \bar{c} 和 \bar{p} 均表示用户 c 和 p 共同评价过的项目的平均值.

2.3 邻居的生成

传统的协同过滤算法在数据稀疏时, 相似度很难计算, 本文提出的策略将相似度和信任度综合考虑, 将 2 个矩阵合并为一个矩阵, 这样合并后的矩阵相对稠密, 合并后矩阵单元的值由相似度和信任度的综合取值, 计算公式如下:

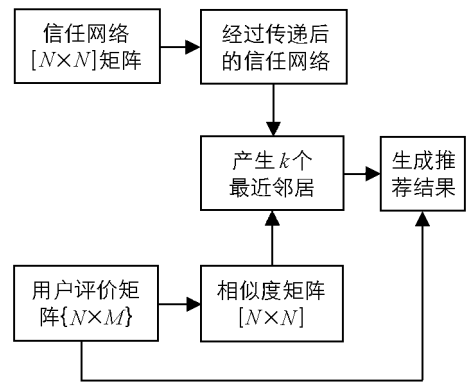


图 2 推荐策略流程图

$$\text{Weight}(i, j) = \begin{cases} \frac{2 * \text{Similarity}(i, j) * \text{Trust}(i, j)}{\text{Similarity}(i, j) + \text{Trust}(i, j)} & \text{if } \text{Similarity}(i, j) > 0 \text{ and } \text{Trust}(i, j) > 0; \\ \text{Trust}(i, j) & \text{if } \text{Similarity}(i, j) = 0 \text{ and } \text{Trust}(i, j) > 0; \\ 0 & \text{if } \text{Trust}(i, j) = 0; \end{cases} \quad (4)$$

其中, $\text{Similarity}(i, j)$ 为用户 i 与用户 j 的相似度, $\text{Trust}(i, j)$ 为用户 i 与用户 j 的信任度, $\text{Weight}(i, j)$ 为合并后的权重. 从合并后的矩阵中选择与目标用户相连且权值最高的 K 个用户作为其邻居.

2.4 产生推荐

采用 Resnick 公式^[2,9,10] 计算目标用户对未知项目的预测评价价值:

$$P(i) = \bar{c} + \frac{\sum_{i=1}^m (p(i) - \bar{p}) * \text{Weight}(c, p)}{\sum_{i=1}^m |\text{Weight}(c, p)|} \quad (5)$$

其中: \bar{c} 为用户 c 的平均评价价值, $p(i)$ 为用户 p 对项目 i 的评价价值, $\text{Weight}(c, p)$ 为推荐的权重, $P(i)$ 为推荐值. 选择推荐值最高且目标用户未评价过的 N 个项目推荐给目标用户.

3 实 验

本文实验在 Intel Pentium IV 2.8G CPU, 内存 256M 的硬件平台下完成, 实验所选用的编程语言是 Java.

3.1 实验数据

实验数据选用的是 EachMovie 提供的标准数据, 选用了其中的 100 个用户, 共有 4 939 条项目评价记录, 稀疏度为 96.97%, 是典型的稀疏数据. 其中信任度的值采用文献[11]中的方法获得.

3.2 评价标准

本文所采用的推荐策略评价标准是平均绝对误差 (V_{MAE})^[3], V_{MAE} 是衡量推荐值与实际评价价值之间误差大小的一种标准, 公式如下:

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

其中, $P(i)$ 为项目 i 的推荐值, $r(i)$ 为项目 i 的真实评价价值, N 为推荐项目总数.

3.3 实验结果

分别采用本文所提出的策略和传统的协同过滤策略进行实验, 流程如图 2 所示, 针对两种策略的对推荐结果分别计算其 V_{MAE} 值, 结果如下:

表 1 实验结果

推荐策略	V_{MAE}
传统的协同过滤推荐策略	0.849 4
基于信任网络的协同过滤推荐策略	0.792 0

实验证明, 基于信任网络的协同过滤推荐策略的推荐准确率有明显提高.

4 结束语

推荐系统是电子商务领域的一项关键技术, 它根据用户的偏好提供个性化的推荐服务, 实现信息过滤. 针对数据矩阵稀疏问题, 在传统协同过滤推荐策略的基础上, 提出了一种在协同过滤中引用信任网络的推荐策略, 通过信任度在信任网络中的传递, 为用户匹配到更多邻居, 将相似度和信任度结合在一起, 使系统推荐的准确率得到明显提高.

参考文献:

- [1] 彭 玉, 陈小平, 徐艺萍. 一种改进的 Item-based 协同过滤推荐算法 [J]. 西南大学学报(自然科学版), 2007, 29(5): 146 – 149.
- [2] Massa P, Bhattacharjee B. Using Trust in Recommender Systems; an Experimental Analysis [A]. Conference on Trust Management [C]. Berlin: Springer, 2005, 2995(5): 221 – 235.
- [3] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems [J]. ACM Transactions on Information System, 2004, 22(1): 5 – 53.
- [4] Massa P, Avesani P. Trust-aware Collaborative Filtering for Recommender Systems [A]. Proceedings of International Conference on Cooperative Information Systems [C]. Berlin: Springer, 2004: 492 – 508.
- [5] Sarwar B M, Karypis G, Konstan J A, et al. Application of Dimensionality Reduction in Recommender System-A Case Study [EB/OL], <http://www.grouplens.org/papers/pdf/webKDD00.pdf>, 02 – 09 – 26.
- [6] 王亚利, 李立新. 基于商务关系网的关联购买策略 [J]. 西南大学学报(自然科学版), 2007, 29(5): 150 – 153.
- [7] Huang Z, Chen H, Zeng D. Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering. ACM Press, 2004, 22 (1): 116 – 142.
- [8] 邓爱林, 朱杨勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法 [J]. 软件学报, 2003, 14(9): 1621 – 1628.
- [9] Resnick P, Iacovou N, Suchak M, et al. Grouplens : An open architecture for collaborative filtering of netnews [A]. ACM Conference on Computer Supported Collaborative Work [C], Chapel Hill: ACM Press, 1994: 175 – 186.
- [10] John O, Barry S. Trust in recommender systems [A]. Proceedings of the 10th International Conference on Intelligent User Interfaces [C]. San Diego: ACM Press, 2005: 167 – 174.
- [11] John O, Barry S. Eliciting trust values from recommendation errors [A]. The 18th International Florida Artificial Intelligence Research Society Conference [C]. Florida : AAAI Press, 2005: 289 – 294.

A Trust Network-based Collaborative Filtering Recommendation Strategy

LU Zhu-bing, TANG Yan

College of Computer and Information Science, Southwest University, Chongqing 400715

Abstract: Collaborative Filtering (CF) is one of the most prevalent recommendation approaches. It provides users with personalized services according to similarity of their preferences. However, the performance of traditional CF method is seriously limited due to the Sparsity problem. A new approach is proposed to deal with this problem. It introduces trust network in traditional CF process. Trust value is propagated through the trust network to match more neighbors for cold start users, and is combined with similarity to generate a compound weight to produce recommendations. Experiment shows that this method is more effective than traditional CF obviously.

Key words: collaborative filtering; recommender system; trust network; trust value

责任编辑 张 枸