

Trust-based New Recommendation Algorithm of Collaborative Filtering Combination

Yajun Liu^{1,2}, Lina Bao², Lisha Gao³

1. School of Computer Science & Engineering, Sanjiang University, Nanjing 210012, China

2. Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, P. R. China

3. Nanjing Power Supply Company, Nanjing 210019, China

Abstract

Aimed at resulting accuracy for common single recommendation technology are not high and there are some limitations, trust relationship in sociology is introduced into personalized news recommendation. This article raises a trust-based new recommendation algorithm of collaborative filtering combination. First, it measures users' trust to recommendation algorithm through the recommendation accuracy; then, in order to raise quality, the recommendation results of user-based and item-based collaborative filtering recommendation algorithm are combined through the trust. The experimental results show that trust-based combining recommendation algorithm have more accurately predict the degree of user interest in news than user-based and item-based collaborative filtering recommendation algorithm and more news of interest to recommend to the user, so it has the effect of better recommended.

Keywords: Trust; News Recommendation; Collaborative Filtering; Combination Recommendation Algorithm

基于信任度的协同过滤组合新闻推荐算法*

刘亚军^{1,2}, 鲍婵娜², 高莉莎³

1.三江学院 计算机科学与工程学院, 江苏 南京 210012

2.东南大学 计算机网络和信息集成教育部重点实验室, 江苏 南京 210096

3.南京供电公司, 江苏 南京 210019

摘 要: 针对常见单一推荐技术产生的推荐结果准确率不高并存在一定局限性问题, 本文将社会学中的信任关系引入到个性化新闻推荐中, 提出了基于信任度的协同过滤组合新闻推荐算法。首先通过推荐的准确率度量用户对推荐算法的信任度, 然后基于信任度将 Item-based 协同过滤推荐算法和 User-based 协同过滤推荐算法的推荐结果进行组合, 以提高推荐的质量。实验结果表明基于信任度的组合推荐算法能更多的向用户推荐感兴趣的新闻, 具有更好的推荐效果。

关键词: 信任度; 协同过滤; 新闻推荐; 组合推荐算法

引言

互联网技术的迅猛发展将人类带入了信息爆炸的时代, 海量数据的产生, 使广大网络新闻的读者受到“信息过载”和“信息迷航”问题的困扰。为了寻找自己感兴趣的新闻信息, 用户往往需要花费大量的时间和精力去选择。目前, 各大新闻类门户网站只能满足主流需求, 不能根据用户自身的喜好提供个性化的服务。然而, 由于用户兴趣的多样性, 人们希望网络能更加智能化, 从而有针对性的为不同用户提供个性化推荐。而这也对各大新闻网站提出了新的要求, 即改变过去对所有用户提供统一界面、同样内容的方式, 针对不同背景、不同兴趣爱好的用户, 提供不同的服务。与此同时, 个性化推荐技术通过分析用户浏览行为, 利用推

*江苏省“十二五”一级学科重点建设学科新闻传播学建设项目系列研究成果

荐算法识别用户兴趣, 过滤他们不感兴趣的内容, 将用户感兴趣的新闻资讯推荐给用户, 从而有效缓解上述问题。

国内已有的新闻资讯推荐, 大都是通过推荐热点的方式来实现推荐, 或者将共同读过某条新闻的其他用户读的新闻推荐给用户^[1], 这种简单的推荐方式, 既不能准确的发现用户的兴趣, 又不能随着用户的访问发现用户兴趣的变化, 因此, 其推荐效果并不理想。另外, 当前大多数推荐系统都需要用户花费额外的时间反馈一些信息, 这给用户带来了一定的负担, 从而降低了用户使用推荐系统的积极性。因此, 为了帮助用户从海量信息中轻松获得感兴趣的新闻资讯, 个性化新闻推荐技术应该从用户的访问行为中获得用户的兴趣, 然后基于用户兴趣向用户推荐个性化的新闻资讯, 并且随用户浏览行为的改变, 及时地发现用户兴趣的变化, 从而动态调整推荐的新闻资讯。因此, 本文提出一种基于信任度的协同过滤组合新闻推荐算法。

1 协同过滤推荐算法

协同过滤推荐算法可分为 Item-based 协同过滤推荐算法和 User-based 协同过滤推荐算法^[2]。Item-based 协同过滤推荐算法是一种基于资源间相似性实现的推荐算法, 因此不能向用户推荐未浏览类别的新闻, 然而新闻间的相似性比较稳定, 可以离线计算后进行保存, 能在一定程度上提高实时推荐的速度。同时, Item-based 协同过滤推荐算法假设用户现在喜欢的新闻与他以前喜欢的新闻都是相似的, 如果用户不满足这个假设, 那么推荐的准确性将不高。User-based 协同过滤推荐算法是一种基于用户相似性实现的推荐算法, 它能发现用户新的兴趣, 但是随着用户数量的增加, 大量用户间相似性的计算将降低实时推荐的效率。同时, User-based 协同过滤推荐算法假设用户会喜欢那些和他有相似兴趣的用户喜欢的新闻, 如果用户不满足这个假设, 那么推荐的准确性也将不高。

由以上分析可知, Item-based 协同过滤推荐算法和 User-based 协同过滤推荐算法在个性化新闻推荐中都有一定的优缺点, 针对这种情况, 本文采用 Item-based 协同过滤推荐算法和 User-based 协同过滤推荐算法相组合的推荐算法, 以充分利用两种推荐算法的优点进行推荐。

2 信任度

现实生活中, 当我们需要朋友建议时, 在兴趣爱好相同的朋友中往往更愿意接受值得信任的朋友建议^[3]。即, 对朋友的信任度将影响我们是否会接受建议。同理, 个性化新闻推荐系统中的用户在接受推荐时也存在同样的问题。因此, 本文将信任度引入到个性化新闻推荐算法中, 通过推荐的准确率度量用户对推荐算法的信任度, 然后基于信任度调整组合推荐的结果。推荐算法推荐的准确率是指该推荐算法向用户 c 推荐并且被接受的新闻占该推荐算法给用户 c 推荐的所有新闻的比例。因此, 信任度 Credit 的计算方法如式 (1) 所示。

$$Credit = Correct = \frac{Accept(c)}{Total(c)} \quad (1)$$

式中, $Accept(c)$ 是推荐算法给用户 c 推荐并且被接受的新闻的数量, $Total(c)$ 是推荐算法给用户 c 推荐的新闻总数。用户 c 接受推荐新闻的数量越多, 用户对推荐算法的信任度将越大, 反之将越小。

3 基于信任度的协同过滤组合新闻推荐算法

用户对推荐算法的信任度越高, 说明用户越容易接受该推荐算法推荐的新闻, 使用该推荐算法对用户进行推荐的准确率将越高, 因此, 本文基于信任度来组合 User-based 协同过滤推荐算法和 Item-based 协同过滤推荐算法, 通过用户对推荐算法的信任度来动态调整相应推荐算法的推荐结果在最终推荐列表中的比例。用户对 User-based 协同过滤推荐算法或 Item-based 协同过滤推荐算法的信任度越大, 该推荐算法给用户推荐的新闻在推荐列表中的比例将越大。基于信任度的协同过滤组合推荐的流程如图 1 所示。

首先基于用户的浏览行为信息分别计算用户对 User-based 推荐算法和 Item-based 推荐算法的信任度; 然后采用 User-based 推荐算法和 Item-based 推荐算法分别生成推荐列表; 接着基于信任度生成最终的推荐列表,

并呈现给用户；用户从推荐列表中选择感兴趣的新闻进行阅读后，又形成了新的浏览行为信息，而这些行为信息又将影响用户对推荐算法的信任度，从而基于信任度的组合推荐算法能根据用户的行为动态调整 User-based 推荐算法和 Item-based 推荐算法推荐的新闻在下一次推荐时的比例。

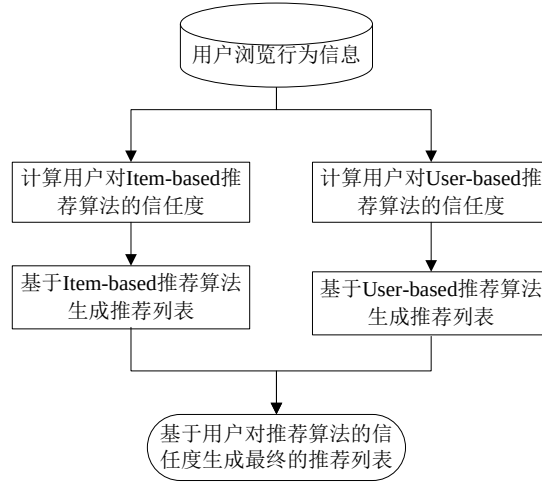


图 1 基于信任度的组合推荐算法流程图

3.1 信任度的计算

用户 c 对 User-based 协同过滤推荐算法的信任度是指 User-based 协同过滤推荐算法给用户 c 推荐并且被用户接受的新闻占该推荐算法给用户 c 推荐的所有新闻的比例。当推荐算法预测用户 c 对新闻的兴趣度与用户 c 实际的兴趣度之差小于一定值 θ 时，即认为用户 c 接受了推荐。因此，用户 c 对 User-based 协同过滤推荐算法的信任度的计算方法如式 (2) 所示，同理，用户 c 对 Item-based 协同过滤推荐算法的信任度的计算方法如式 (3) 所示。信任度的取值可随着用户的反馈行为动态调整。

$$Credit_{user} = \frac{\sum_{i \in R_{user,c}} f_{user}(i)}{|R_{user,c}|} \quad (2)$$

$$Credit_{item} = \frac{\sum_{i \in R_{item,c}} f_{item}(i)}{|R_{item,c}|} \quad (3)$$

式 (2) 中， $R_{user,c}$ 是 User-based 协同过滤推荐算法给用户 c 推荐的所有新闻的集合， $f_{user}(i)$ 的定义如式 (4) 所示；式 (3) 中， $R_{item,c}$ 是 Item-based 协同过滤推荐算法给用户 c 推荐的所有新闻的集合， $f_{item}(i)$ 的定义如式 (5) 所示。

$$f_{user}(i) = \begin{cases} 1, & |Reco_{user,i} - Value_i| < \theta \\ 0, & other \end{cases} \quad (4)$$

$$f_{item}(i) = \begin{cases} 1, & |Reco_{item,i} - Value_i| < \theta \\ 0, & other \end{cases} \quad (5)$$

式 (4) 中， $Reco_{user,i}$ 是 User-based 协同过滤推荐算法预测的用户 c 对新闻 i 的兴趣度；式 (5) 中， $Reco_{item,i}$ 是 Item-based 协同过滤推荐算法预测的用户 c 对新闻 i 的兴趣度。 $Value_i$ 是用户 c 对新闻 i 的实际兴趣度。 θ 是兴趣度预测误差阈值，当推荐算法预测的兴趣度与用户实际的兴趣度之差小于 θ 时，该推荐才是被用户接受的。

3.2 基于 Item-based 推荐算法生成推荐列表

将用户已阅读新闻和目标新闻表示成基于概率的向量空间模型^[4]；然后计算用户已阅读新闻与目标新闻的相似性，取相似性较高的新闻构成目标新闻的最近新闻集合；接着通过用户对最近新闻集中新闻的兴趣

度预测用户对目标新闻的兴趣度；最后将所有目标新闻按预测得到的兴趣度大小进行排序，生成推荐列表 L_{item} 。

3.3 基于 User-based 推荐算法生成推荐列表

首先根据基于分类的用户兴趣建模方法表示用户的兴趣；然后计算用户之间的相似性，取相似度较高的用户构成最近邻居集合^[5]；接着通过最近邻居集合中的用户对目标新闻的兴趣度，预测当前用户对目标新闻的兴趣度；最后将所有目标新闻按预测得到的兴趣度大小进行排序，生成推荐列表 L_{user} 。

3.4 基于信任度调整推荐结果

通过 User-based 协同过滤推荐算法和 Item-based 协同过滤推荐算法所生成的推荐列表 L_{user} 和 L_{item} ，然后基于用户对 User-based 协同过滤推荐算法和 Item-based 协同过滤推荐算法的信任度 $Credit_{user}$ 和 $Credit_{item}$ 来调整两种推荐算法推荐的新闻在最终推荐列表中的比例，调整方法如式（6），（7）所示。

$$Num_{user} = \frac{Credit_{user}}{Credit_{user} + Credit_{item}} \times |L| \quad (6)$$

$$Num_{item} = \frac{Credit_{item}}{Credit_{user} + Credit_{item}} \times |L| \quad (7)$$

式中， $|L|$ 是最终推荐列表的大小， Num_{user} ， Num_{item} 分别是 User-based 协同过滤推荐算法和 Item-based 协同过滤推荐算法推荐的新闻在最终推荐列表中的数量，用户对推荐算法的信任度越高，该推荐算法给用户推荐新闻的数量将越多。

4 实验与分析

基于信任度的组合推荐算法，主要是基于新闻间的相似性和用户间的相似性实现推荐。由于新闻之间的相似性比较稳定，可以离线计算后进行保存，因此在线阶段只需计算用户之间的相似度以及用户对推荐算法的信任度，而这两者的计算与用户数量成正比，因此，假设 N 为用户的数量，本文提出的基于信任度的协同过滤组合推荐算法的复杂度为 $O(N)$ 。

为了验证基于信任度的协同过滤组合推荐算法的有效性，实验在基于算法实现的新闻推荐系统之上，模拟了若干用户，采用平均绝对误差(MAE)和推荐覆盖率(coverage)对推荐结果进行评价。平均绝对误差通过推荐算法预测的评分与用户实际的评分之间的偏差来度量预测的准确率，推荐覆盖率通过推荐系统为用户推荐的资源对用户兴趣的覆盖范围来度量预测的质量^[6]。

首先，通过实验对兴趣度预测误差 θ 阈值的值进行设置，实验在多次推荐中动态调整 θ 的值，得到如图 2、图 3 所示的实验结果。

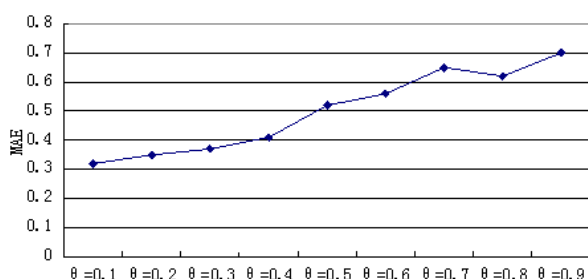


图 2 平均绝对误差随 θ 变化图

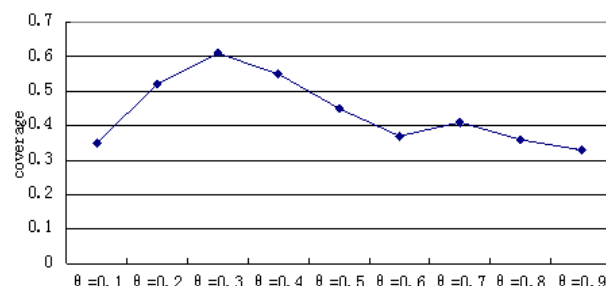


图 3 推荐覆盖率随 θ 变化图

从图 2 可以看出， θ 取值越小时推荐准确率越高，这说明预测兴趣度误差阈值越小越有利于提高推荐的准确率。从图 3 可以看出，当 θ 取值过小或过大时推荐覆盖率都不高，当 θ 取 0.3 时具有最高的推荐覆盖率，此时的推荐准确率也较高，虽然当 θ 取 0.2 和 0.1 时的推荐准确率高于取 0.3 时，但他们的差别很小，而 θ 取

0.3 的推荐覆盖率明显高于取 0.1 和 0.2 时。综合上述分析, 预测兴趣度误差阈值 θ 取 0.3。

为了验证本文提出的组合推荐算法的有效性, 实验中分别采用 Item-based 协同过滤推荐算法(IB), User-based 协同过滤推荐算法(UB)以及基于信任度的组合推荐算法(CUI), 进行多次推荐, 由推荐结果分别得到了图 4、图 5 所示的平均绝对误差对比图和推荐覆盖率对比图。

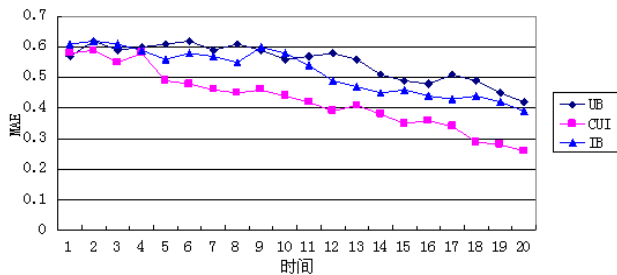


图 4 各推荐算法平均绝对误差对比图

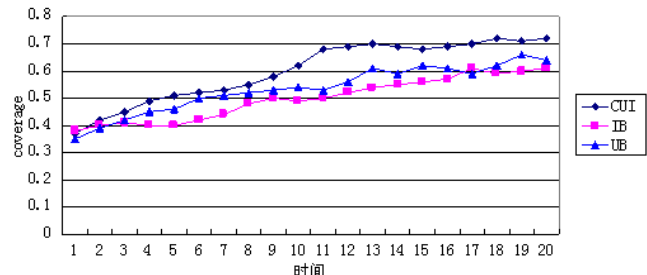


图 5 各推荐算法推荐覆盖率对比图

从各推荐算法平均绝对误差对比图 4 可以看出, 基于信任度的组合推荐算法相比于 Item-based 和 User-based 协同过滤推荐算法获得了较小的平均绝对误差, 并且随着时间的推移, 算法表现了良好的性能, 即预测准确率得到了一定的提高, 说明通过信任度调整推荐算法在推荐结果中的比例能在一定程度上提高预测用户兴趣的准确率, 从而推荐系统能更准确的识别用户的兴趣, 进而更好的向用户推荐感兴趣的新闻。

从各推荐算法推荐覆盖率对比图 5 可以看出, 基于信任度的组合推荐算法相比于单个 Item-based 或 User-based 协同过滤推荐算法获得了较大的推荐覆盖率, 即基于信任度的组合推荐算法推荐的新闻比 Item-based 和 User-based 协同过滤推荐算法推荐的新闻, 占用户感兴趣新闻的比例大, 说明组合推荐算法能在一定程度上提高推荐的覆盖率, 从而推荐系统能向用户推荐更多的感兴趣新闻。

5 结束语

现有的新闻推荐系统大都不能随着用户的浏览行为及时地捕捉到用户兴趣的变化, 因此推荐的准确率不高。为了提高个性化新闻推荐的质量, 本文提出了基于信任度的协同过滤组合新闻推荐算法, 实验结果证明, 该算法比 Item-based 和 User-based 协同过滤推荐算法能更准确地预测用户对新闻的兴趣度, 更多的向用户推荐感兴趣的新闻, 因此具有更好的推荐效果。

REFERENCES

- [1] 新浪新闻中心.<http://news.sina.com.cn>
- [2] 曾春, 邢春晓, 周立柱. 个性化服务技术综述[J]. 软件学报, 2002, 13(10): 1952-1961
- [3] 张迎峰. 面向数字图书馆的个性化推荐算法研究[D]. 合肥: 中国科学技术大学, 2011
- [4] 蒲筱哥. 自动文本分类方法研究述评[J]. 情报科学, 2008, 26(3): 469-475
- [5] 李聪, 梁昌勇, 马丽. 基于领域最近邻的协同过滤推荐算法[J]. 计算机研究与发展, 2008, 45(9): 1532-1538
- [6] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 5-53

【作者简介】



刘亚军 (1953-), 女, 汉, 三江学院、东南大学, 教授。研究方向: 计算机应用技术。毕业于东南大学。

Email: yjliu@seu.edu.cn

鲍婵娜 (1987-), 女, 汉, 东南大学, 硕士研究生。研究方向: 计算机应用技术, 就读于东南大学。

Email: 306189888@qq.com

高莉莎 (1982-), 女, 汉, 南京供电公司, 硕士, 高级工程师。研究方向: 计算机应用技术, 毕业于东南大学。

Email: gaolisha@js.sgcc.com.cn