Computer Science and Application 计算机科学与应用, 2011, 1, 108-111
http://dx.doi.org/10.12677/csa.2011.13022 Published Online December 2011 (http://www.hanspub.org/journal/csa)

Optimized Cooperative Filtering Algorithm#

Ruixin Ma*, Fancheng Meng, Hanyang Wang
School of Software, Dalian University of Technology, Dalian
Email: teacher_mrx@126.com
Received: Oct. 9th, 2011; revised: Nov. 1st, 2011; accepted: Nov. 18th, 2011.

Abstract: Cooperative Filtering algorithm is one of the most successful technologies in the application of personalized recommendation system. However, the users' identities and interests are very complicated in the normal network, which results in the inaccuracy of the recommendation. However, the appearance of social network provides users with a relatively green and safe communication platform. This paper in terms of the problems in the personalized recommendation system of social network comes up with the trust-based nearest neighbor algorithm which optimizes the traditional NN algorithm and provides recommendation by analyzing trust between the social network users. The experimental results show that the modified algorithm greatly improves the recommendation accuracy and raises the users' degree of satisfaction.

Keywords: Social Network; Trust; Personalized Recommendation; The Nearest Neighbor

优化的协同过滤推荐算法#

马瑞新*，孟繁成，王涵杨
大连理工大学软件学院，大连
Email: teacher_mrx@126.com
收稿日期：2011 年 10 月 9 日；修回日期：2011 年 11 月 1 日；录用日期：2011 年 11 月 18 日

摘　要：基于最近邻的协同过滤算法是个性化推荐中应用的最为成功的算法，然而在普通网络中，用户身份兴趣等信息缺乏可靠性，导致该算法推荐不够精确。社会网络是彼此拥有足够的感情后建立起来的一种人际关系网络，它的出现为广大用户提供了一个相对安全的绿色交流通道。本文针对社会网络中个性化推荐系统存在的问题，提出了一种在社会网络下基于信任度的最近邻居集合算法的优化，该算法通过分析社会网络中用户的信任度给出个性化推荐。实验表明，该算法在社会网络中的应用提高了个性化推荐的准确度，增加了用户的满意程度。

关键词：社会网络；信任度；个性化推荐；最近邻居

1. 引言

个性化推荐系统是为解决互联网上信息过载问题而提出的一种智能代理系统，它能从大量信息中向用户推荐出符合其兴趣偏好或需求的资源[1]。在个性化推荐系统中，应用最成功的算法是最近邻居的协同过滤算法。协同过滤算法主要有基于用户的和基于项目的 2 种算法[2]。基本思想是通过计算目标用户与各个基本用户对项目评分之间的相似性，搜索目标用户的最近邻居，然后由最近邻居的评分数据向目标用户产生推荐，即目标用户对未评分项目的评分可以通过最近邻居对该项目评分的加权平均值进行逼近，从而产生推荐[3]。

为了寻找目标用户的最近邻居集合，需要度量用户之间的相关性。在传统网络中，由于用户兴趣偏好或需求的资源往往存在巨大的差异，这样找出来的最近邻居集合不够准确。社会网络是由一定数量的用户在在网络上经过一段时间的交流、讨论，彼此拥有足够的感情后建立起来的一种人际关系网络[4]，它涵盖了以人类社交为核心的所有的网络服务形式，是一个真正能够相互交流、相互沟通、相互参与的平台，因

#基金项目：中央高校基本科研业务费专项资金资助。

此，社区用户之间彼此更加依赖，社区用户在做出某项决定之前往往会咨询社区内的其它用户，寻求推荐意见。社会网络由于其相对真实的资料注册门槛，在很大程度上纯洁了注册用户的身份来源，为广大用户提供了一条更加绿色、安全的交友渠道，极大提高了网络人际传播的效率，优化了社会网络交往的效果。在社会网络中，信任度成为一个重要的因素，本文针对信任度对最近邻算法进行优化，使之能在社会网络中更好发挥推荐作用。

2. 基于最近邻的协同过滤推荐算法

传统的计算最近邻居集合的方法是相关相似性度量。用户的相似性的度量主要包括三种：余弦相似性、Pearson 相关相似性、改进的余弦相似[5]。

1) 余弦相似性：用户评分被看成是 n 维项目空间上的向量，用户间的相似性通过向量间的余弦夹角来度量。设用户 u 和 v 在 n 维项目空间上的评分分别为向量 m,n，则用户 u 和用户 v 之间的相似性 $sim(u,v)$ 为：

$$sim(u,v)=\cos(m,n)=\frac{m\cdot n}{|m|\cdot|n|} \tag{1}$$

2) 似性：设经用户 u 和用户 v 共同评分过的项目用 $I_{u,v}$ 表示，则用户 u 和用户 v 之间的相似性 $sim(u,v)$ 经过 Pearson 相关系数度量：

$$sim(u,v)=\frac{\sum i\in I_{u,v}\left(R_{u,i}-\overline{R_u}\right)\left(R_{v,i}-\overline{R_v}\right)}{\sqrt{\sum i\in I_{u,v}\left(R_{u,i}-\overline{R_u}\right)}\sqrt{\sum i\in I_{u,v}\left(R_{v,i}-\overline{R_v}\right)}} \tag{2}$$

其中，$R_{u,i}$ 表示用户 u 对项目 i 的评分，$\overline{R_u}$，$\overline{R_v}$ 分别表示用户 u 和用户 v 对所有属于 $I_{u,v}$ 中项目的平均评分。

3) 修正的余弦相似性：由于在余弦相似性度量方法中没有考虑不同用户的评分尺度问题，修正的余弦相似性通过减去用户对项目的平均评分来改善上述缺陷。设经用户 u 和用户 v 共同评分的项目集合用 $I_{u,v}$ 表示，I_u 和 I_v 分别表示经用户 u 和用户 v 评分过的项目集合，则用户 u 和用户 v 之间的相似性 $sim(u,v)$ 为：

$$sim(u,v)=\frac{\sum i\in I_{u,v}\left(R_{u,i}-\overline{R_u}\right)\left(R_{v,i}-\overline{R_v}\right)}{\sqrt{\sum i\in I_u\left(R_{u,i}-\overline{R_u}\right)}\sqrt{\sum i\in I_v\left(R_{v,i}-\overline{R_v}\right)}} \tag{3}$$

其中，$R_{u,i}$ 表示用户 u 对项目 i 的评分，$R_{u,i}$，$\overline{R_v}$ 分别表示用户 u 和用户 v 对项目的平均评分。

以上 3 种方法均为基于向量的相似度量计算方式，进行对象属性之间的严格匹配。修正的余弦相似性方法在余弦相似性基础上，减去了用户对项目的平均评分，如果用户间所有的评分项目均为共同评分项目，那么相关相似性与修改的余弦相似性是相同的。

通过以上的算法分析可知，相似度是整个最近邻居计算过程中的唯一决定因素，产生的推荐是基于用户最近邻评价的平均加权值的 Top 推荐，权重是最近邻之间的用户相似度。

实际上，在社会网络中，即使社区用户的兴趣模型不是非常相似，但由于社区用户在日常生活、学习和工作上的交流和沟通的逐渐加深，在交往过程中不断地接受影响和被影响，彼此之间会形成一种彼此信任的关系。这种依赖性使得社区用户在进行某项活动之前会主动询问他所信任的其他用户的意见，然后进行选择性采纳他人意见。因此，传统的单纯用户相似性度量并不适用于社会网络下的最近邻居集合计算。社会网络中的用户之所以会采纳其他用户的意见，是基于彼此之间的信任。社区用户在信息交互的过程中，逐渐形成这样的认识：接受社区中合作伙伴的意见会使自己受益(专家或者有经验的人的指导总是强于自己没有头绪的尝试)，即关系交互用户愿意去依赖他所信任的同伴。由此可知，信任关系在社会网络下研究用户行为的相互影响过程中，有着不可替代的作用。

3. 优化的协同过滤推荐算法

针对于社会网络中用户之间存在的关系，本文提出一种基于信任度的最近邻居集合算法。信任元 T(u, v) 表示社区用户 u 对社区用户 v 的信任程度，0 < T < 1。在社会网络中，用户之间的信赖关系可以分为两种：静态信任和动态信任。

静态信任(static trust)是指社区用户设定的自己与其它社区用户之间的信任关系，它是一种显式的信任表示程度，本文中用用户优先级来表示。社会网络用户的交流、分享的过程中，彼此之间的关系会逐渐变得“明朗化”，因此能够根据自己的喜好、评判准则，对于其他的社区用户设置不同的身份。以校内网为例，社区用户的身份可以是特别好友、家人亲属、大学同学等等。社会网络对于不同的身份设置不同的优先级

别，优先级别越高，该身份受到的信任程度就越大。

动态信任(dynamic trust)是指在信息咨询与反馈过程中建立的用户之间的信任程度，它是一个动态的矢量，是系统记录的用户之间隐式的信任关系。在社会网络中，社会网络用户发起信息咨询的时候，会遇到两种情况。有的用户很热心，愿意对他人的咨询做出回应，这样的用户会逐步赢得信息咨询者的信任，在社交圈中建立自己的威信；有的用户对待他人的咨询很冷淡，不理会他人的咨询，这样的用户会逐步丧失信息咨询者的信任。

设 d$T(u, v)$为社区用户 u 对社区用户 v 的动态信任程度，$f_i(u, v)$表示用户 u 第 i 次向用户 v 发出的信息咨询，$b_i(u, v)$表述用户 v 对用户 u 的信息咨询做出的反映，若 v 对 u 进行了回应，则 $b_i(u, v)$取值为 1，反之为 0。此外，用逻辑斯帝函数对不同时间段的信息交互进行时间加权，用户之间最近的信息交互被赋予较大的权重，过去的信息交互则赋予较小的权重。则动态信任可用下式计算：

$$dT(u,v)=\frac{\sum_{i=1}^{n}\text{logistic}\left(t_{u,v}^{i}\right)\times b_i(u,v)}{\sum_{i=1}^{n}\text{logistic}\left(t_{u,v}^{i}\right)\times f_i(u,v)} \tag{4}$$

$$\text{logistic}\left(t_{u,v}^{i}\right)=\frac{1}{1+e^{-t_{u,v}^{i}}} \quad -1<t_{u,v}^{i}<1 \tag{5}$$

$$0<\text{logistic}\left(t_{u,v}^{i}\right)<1$$

$\text{logistic}\left(t_{u,v}^{i}\right)$ 是单调递增函数，输出值随着 t 一直增加并且权值始终保持在(0, 1)范围内。实验前，预先应用标准化变量将时间变量 t 的范围转化为–1~1，在此范围中，它们的行为几乎呈线性，且时间变量 t 的微小变化会导致权重的微小变化，从而能够更加精确的追踪用户之间的关系变化。

信任度是用户之间动态信任与静态信任的调和权重，计算公式如下所示：

$$T(u,v)=\left(sT(u,v)+\frac{2\times sT(u,v)\times dT(u,v)}{sT(u,v)+dT(u,v)}\right)\Big/2 \tag{6}$$

实验前，预先应用标准化变量将 $sT(u, v)$和 $dT(u, v)$的范围转化为 0~1。该式考虑到社会网络的特点，即社会网络本身就是建立在一定的信任度的基础上的。

4. 实验结果分析对比

本文采用山东某大学校内网的用户数据库作为实验数据集，它由该大学的 1200 多名校内网用户的基本数据库和关系数据库组成，每个用户至少有 100 名校内好友。

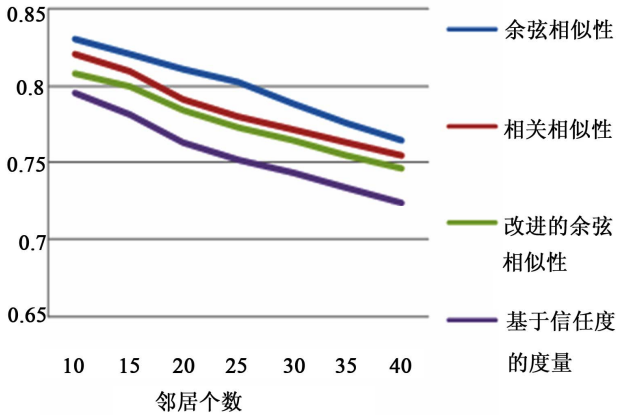

Figure 1. The MAE of different recommended strategies
图 1. 不同推荐策略的 MAE

本文采用平均误差标准(MAE)进行度量，它主要用于度量真实值与预测值之间的偏差，偏差越小，度量的精确度越高，用户之间关系度量越准确。

$$\text{MAE}=\frac{\sum_{i,j\in\text{friends}}\left(T(i,j)-\overline{T(i,j)}\right)}{\text{No}_{\text{friends}}} \tag{7}$$

$T(i, j)$为系统预测用户 i 对用户 j 的信任度，$\overline{T(i,j)}$ 为用户 i 对用户 j 的实际信任度，用户 i 与用户 j 为好友关系。

为了验证本文提出的计算最近邻居集合方法的有效性，与传统的用户相关相似性度量进行了试验比较，实验结果如图 1 所示。

为了验证本文提出的计算最邻近居集合方法的有效性，设置不同的最近邻居的个数，查看不同数目的邻居集合对方法精确度的影响。由图 1 可知，对于不同的最近邻居数目，基于信任度的最近邻居集合计算方法均要由于传统的用户相关相似性度量。由此可知，本文提出的计算最近邻居集合的方法精确度更高。

5. 结论

本文提出信任是社会网络中判断用户之间关系密切程度的标准，并且为此构建了切实合理的信任模型。实验分析证明，把信任度引入到社会网络下的最近邻居计算中是合理可行的，提高了最近邻居集合的准确性。

参考文献 (References)

[1] 卢志国, 尹雪, 蒋丽丽. 信息共享环境下的社交网络[J]. 数字图书馆论坛, 2009, 2: 31-35.

[2] B. Sarwar, G. Karypis, J. Konstan, et al. Item-based collaborative filtering recommendation algorithms. Hong Kong: Proceeding of the 10th International World Wide Web Conference, 2001: 285-295.

[3] 张光卫, 李德毅, 李鹏. 基于云模型的协同过滤算法的研究[J]. 软件学报, 2007, 18(10): 2403-2410.

[4] Y. Ding, X. li. Time collaborative filtering. Bremen: Proceeding of Conference on Information and Knowledge Management, 2005: 485-492.

[5] J. A. Michael, B. G. S. Linoff. 数据挖掘技术[M]. 北京: 机械工业, 2008.

