自动建立信任的防攻击推荐算法研究

黄世平[1], 黄 晋[2,3], 陈 健[4], 汤 庸[3]
(1. 中山大学信息科学与技术学院, 广东广州 510006; 2. 深圳移动互联网应用中间件技术工程实验室, 广东深圳 518060;
3. 华南师范大学计算机学院, 广东广州 510631; 4. 华南理工大学软件学院, 广东广州 510006)

摘 要: 随着互联网中信息资源的日益增多, 个性化推荐技术作为缓解"信息过载"的有效手段, 得到了越来越多的研究者的关注. 由于互联网天然的开放性, 在商业利益的驱动下, 部分恶意用户通过伪造虚假数据来影响系统的推荐结果, 从而达到盈利的目的. 本文提出一个自动建立信任的防攻击推荐算法, 在考虑了用户评分相似性的基础上, 引入适当的信任机制, 通过为目标用户动态建立和维护有限数量的信任对象来获得可靠的推荐. 大量基于真实数据集的实验表明, 提出的算法能大大提高推荐系统的鲁棒性和可靠性, 并在一定程度上提高了推荐的精准度.

关键词: 推荐系统; 用户信任; 恶意攻击
中图分类号: TN911. 23 文献标识码: A 文章编号: 0372-2112 (2013)02-0382-06
电子学报 **URL**: http://www.ejournal.org.cn **DOI**: 10.3969/j.issn.0372-2112.2013.02.027

Anti-Attack Recommender Algorithm Based on Automatic Trust Establishment

HUANG Shi-ping[1], HUANG Jin[2,3], CHEN Jian[4], TANG Yong[3]
(1. *School of Information Science and Technology, Sun Yat-sen University, Guangzhou, Guangdong* 510006, *China*;
2. *Shenzhen Engineering Laboratory for Mobile Internet Application Middleware Technology, Shenzhen, Guangdong* 518060, *China*;
3. *School of Computer, South China Normal University, Guangzhou, Guangdong* 510631, *China*;
4. *School of Software Engineering, South China University of Technology, Guangzhou, Guangdong* 510006, *China*)

Abstract: As the information resources available on the Internet are booming nowadays, personalized recommendation technique, which is an effective approach to ameliorate information overloading, has increasingly received attentions from researchers. Due to the native open nature of the Internet and driven by commercial motives, some malicious users attempt to influence the recommendation result via faking data, hoping to gain profits by manipulating recommendation. This paper proposes an anti-attack recommendation algorithm based on automatic trust establishment. Considering the similarities between user ratings, the proposed algorithm introduces a trust mechanism to obtain reliable recommendations through dynamically constructing and maintaining trusted references for users. Enormous experimental results obtained from real datasets reveal that the proposed algorithm could significantly improve both robustness and reliability of recommendation system, and meanwhile enhance the accuracy of recommendation to some extent.

Key words: recommender system; user trust; malicious attack

1 引言

随着互联网应用逐渐遍布商业、社会、生活等各个方面, 其中"信息过载"问题慢慢成为一个不可忽视的困扰. 个性化推荐系统捕捉用户在互联网应用中体现的兴趣爱好, 为其推荐其有可能感兴趣的实体, 包括新闻、商品、朋友、博客、评论、社区等等, 从而实现了信息服务由被动到主动的跨越, 是目前学术界和工业界的一个研究热点[1~3].

但由于互联网天然的开放性, 使部分竞争者可以通过伪造用户的偏好数据来影响系统对其他用户的推荐结果, 从而达到提高自己的商业利益或损害对手的商业利益的目的. 随着在线购物和互联网上资源共享等应用的兴起, 面对推荐攻击的挑战, 如何在不改变用户使用习惯的前提下采用普遍适用的机制增强系统的鲁棒性, 是一个亟待解决的问题.

收稿日期: 2012-07-16; 修回日期: 2012-09-30
基金项目: 国家自然科学基金重点项目(No. 60736020); 国家自然科学基金(No. 60970044, No. 61272067, No. 61272065); 广东省自然科学基金(No. S2012010009311); 广东省科技项目(No. 2011A091000036, No. 2011168005, No. 2011B080100031); 华南理工大学中央高校基本科研重点项目(No. 2012ZZ0088)

目前已有研究尝试将社会网络中的信任关系结合到传统的推荐系统中, 以提高推荐的精准度. 虽然这种方式在一定程度上可降低被攻击的可能, 但同时也增加了用户的负担, 用户需显式的给出对其他用户的信任值, 并自行维护和调整自己的信任列表. 本文提出一个基于信任分层的防攻击推荐算法, 从用户的历史购买记录中分析和挖掘用户之间的信任关系, 并根据用户新的购买行为动态的调整信任关系, 在考虑了用户评分相似性的基础上, 通过为目标用户维护有限数量的信任对象来获得可靠的推荐.

2 相关工作

用户概要文件是指推荐系统中记录的关于用户兴趣的个人数据, 如个人购买记录及评分等. 由于推荐系统依赖用户的概要文件来计算相似度, 因此恶意用户可以通过伪造虚假的用户概要文件来尽量贴近其他用户的兴趣爱好, 从而进入到目标用户的最近邻列表, 对推荐产生影响. 推荐攻击的一般形式如图 1[4] 所示, 其中 I_S 是由攻击者选定的具有某一特性的项目集合, I_F 是随机选择的用于填充的一组项目集合, I_Φ 是一组未评分项目集合和某个目标项目 i_t.

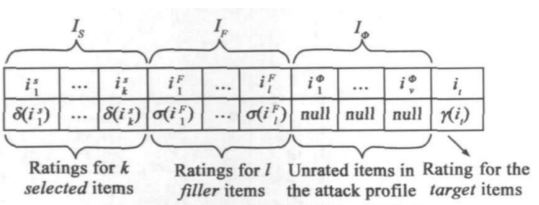


图1 伪造用户偏好数据的一般形式

不同的攻击模型的区别就在于选择 I_S 的方式、I_F 的填充比例以及对目标项目 i_t 的评分. 根据攻击的目的, 攻击模型大体上分为两种: 增加目标项目被推荐机会的推动攻击(push attack)和降低目标项目被推荐机会的核攻击(nuke attack).

根据文献[4] 得到的结论, 各种攻击模型对推荐产生的影响会越来越大, 这就为研究防攻击的推荐算法提出了要求.

目前有两种常见的对抗攻击的形式. 一种是攻击检测, 一般基于数理统计, 如文献[4] 中提出的从用户偏好数据的长度、对某一项目评分的偏离、背离平均度和与最近邻的相似程度等方面来衡量和判断, 该检测方法虽然比较直观, 但检测比较复杂, 全局设定的指标阈值会将一些真实用户误认为是攻击者, 因此检测准确率并不十分理想. 为了降低计算量, Chirita 等人[5] 提出了同名模型, 只计算背离平均度和攻击可能度, 但随后被证明了该模型对低填充(I_F 较少)的攻击检测效果不明显. Burke 等人[6] 进一步提出了分块检测模型, 采用注入平均目标差异和相符权重度作为检测指标, 但也被验证了在实际中对高攻击规模的攻击检测效果也不太好. 正如文献[7] 中提到的, 当攻击记录仿真程度提高, 基于检测方法的查准率与查全率都不理想, 可能将偏好与大众偏好不同的用户错误地识别为攻击用户, 又可能会造成个性化信息丢失的后果. 另外一种是防御攻击, 如利用社会网络中已建立的用户关系, 从而将攻击者尽可能地排除在外. Ma 等人[8] 提出将社会信任作为一个推荐的约束, 使用概率因子分析的方式来结合用户与其所信任的用户间的信任关系来进行推荐. 文献[9] 在分布式社会网络环境中提出 SybilGuard 协议, 通过分析攻击者和正常用户在社会网络中形态的不同来区分二者, 进而限制可能的攻击者与其他用户创建联系的数目, 来减少攻击造成的影响.

综上分析, 目前基于攻击检查的方法并不能完全杜绝伪造虚假用户概要的文件的攻击, 而基于社会网络的防御攻击方法需要维持所有用户间的信任值, 代价相对较大, 且可行性并不高.

3 基于信任分层的推荐系统

协作过滤推荐框架中, 用户的事务数据库被表示为 m 个用户的集合 $U=\{u_1, u_2, \cdots, u_m\}$ 和 n 个项目的集合 $I=\{i_1, i_2 \cdots i_n\}$.

传统的协作过滤推荐系统, 是根据用户的相似性来推荐资源的. 该算法基于这样一个假设: "如果用户对某些页面/项目体现出相似的兴趣(如对页面相似的访问模式或对项目相似的评分), 那么他们对其它的页面/项目也具有较为相似的爱好". 因此该算法将每个用户的兴趣建模为一个评分向量, 通过计算向量之间的相似性为当前在线用户寻找 k 个最相似的邻居. 针对某个特定的, 当前用户尚未评分或浏览的项目, 根据邻居们感兴趣程度对其作出预测, 或是直接将邻居们最感兴趣的 N 个项目推荐给当前用户(Top-N Recommendation).

由于过度依赖历史评分数据, 传统的协作过滤推荐系统容易受到系统中攻击数据的影响; 而已有的基于信任关系的推荐方法只考虑目标用户 k 个最相似的邻居的信任度的调整, 当遇到攻击时, 不足以给用户提供准确的推荐. 因此, 本文提出了基于信任分层的防攻击推荐算法.

定义 **1** (信任值)　若用户 u_i 信任用户 u_j, 则记为 $u_i \xrightarrow{tru_{i\to j}} u_j$, 其中信任值 $tru_{i\to j}$ 表示信任程度, 取值范围在 [0, 1] 之间, 值越大表示信任程度越高. 此信任值不依赖任何外部信息, 而是赋予初始值 t_0 后由系统根据历史评分记录动态调整.

定义 **2**（相似度） 若用户 u_i 与用户 u_j 相似，则记为 $u_i \xleftrightarrow{sim_{i,j}} u_j$，其中相似度 $sim_{i,j}$ 表示相似程度，取值范围在[0，1] 之间，值越大表示相似程度越高. 该值在本文中通过标准的余弦相似度来衡量.

定义 **3**（推荐权重） 用户 u_i 对用户 u_j 间的推荐权重 $w_{i,j}=tru_{i,j}\times sim_{i,j}$，用于衡量用户 u_i 在为用户 u_j 产生推荐结果时，其意见的重要性，重要性越高取值越大.

定义 **4**（信任层） 对于目标用户 u_t，其信任层记为 H_t^R，记录了前 k 个推荐权重最大的用户集合（对 u_t 而言）. 该集合将用于为 u_t 产生推荐结果.

定义 **5**（缓冲层） 对于目标用户 u_t，其缓冲层记为 H_t^B，记录了次大的 k 个推荐权重最大的用户集合（对 u_t 而言）. 该集合不直接用于为 u_t 产生推荐结果.

具体的，系统的推荐流程如下，其中 $|H_t^R|$，$|H_t^B|$ 为 k，初始化信任程度为 t_0：

（1）初始化，系统根据推荐权重 $w_{i,j}=tru_{i,j}\times sim_{i,j}$，$u_i$，$u_j\in U$（所有用户），为每个用户 u_t 构造 H_t^R 与 H_t^B.

（2）对于每个用户 u_t，用 H_t^R 中的用户评分为其计算未评分项目的预测评分，并将评分最高的项目推荐给用户 u_t，对项目 i_m 的个性化预测评分 $r_{t,m}$ 由式（1）得到：

$$r_{t,m}=\bar{r}_t+\frac{\sum_{u_v\in H_t^R} w_{t,v}\times\left(r_{v,m}-\bar{r}_v\right)}{\sum_{u_v\in H_t^R}|w_{t,v}|} \tag{1}$$

其中 $\bar{r}_t$ 与 $\bar{r}_v$ 分别为用户 u_t 与 u_v 对所有项目的评分平均值，$r_{v,m}$ 为用户 u_v 对项目 i_m 的评分值.

（3）当用户接受推荐并使用项目后，用户 u_t 将对项目 i_m 给出真正评分 $r'_{t,m}$，系统会根据 $r'_{t,m}$ 和 $r_{v,m}$ 的差异调整对用户 u_v 的信任值 $tru_{t\to v}$，其中 $u_v\in H_t^R\cup H_t^B$，并重新构造 $H_t'^R$，$H_t'^B$，然后返回步骤（2），具体如下：

（*a*）更新信任值. 若 $r'_{t,m}=r_{v,m}$，则说明用户 u_v 的推荐符合用户 u_t 的兴趣爱好，u_t 对 u_v 的信任程度 $tru_{t\to v}$ 应该增加，反之应该减少. 参考文献[7]，我们定义评分误差

$$Err_{t,v,m}=\frac{|r'_{t,m}-r_{v,m}|}{(r_{\max}-r_{\min})/2} \tag{2}$$

其中 $r_{\max}$ 和 $r_{\min}$ 分别是系统运行的最大与最小评分值. 给定参数 φ，设 $R_{t,v,m}=\varphi-Err_{t,v,m}$，调整前的信任值为 $tru_{t\to v}$，调整后的信任值为 $tru'_{t\to v}$，则有：

$$tru'_{t\to v}=(1-R_{t,v,m})\times tru_{t\to v}+R_{t,v,m} \tag{3}$$

即当 $R_{t,v,m}>0$ 时信任程度 $tru'_{t\to v}$ 增加，当 $R_{t,v,m}<0$ 负面反馈时信任程度 $tru'_{t\to v}$ 减少. 若 $Err_{t,v,m}=\varphi$ 则信任程度 $tru'_{t\to v}$ 不变. 显然的，参数 φ 取值越小，系统对误差的定义越严格. 注意到由式（3）计算得到的信任值 $tru'_{t\to v}$ 可能会超出[0，1] 范围，我们约束在[0，1] 范围内.

（*b*）更新层次. 当信任程度调整后，对于所有的 $u_v\in H_t^R\cup H_t^B$，推荐权重 $w_{t,v}$ 也需要根据定义 3 重新计算. 根据重新计算的 $w'_{t,v}$ 调整 H_t^R，H_t^B，始终使推荐权重最高的前 k 个用户保持在 H_t^R 中.

（4）随着用户新评分项目的增多和新用户的加入，系统运行一段时间 t 后，对于用户 u_i，$u_j\in U$（所有用户），相似程度 $sim_{i,j}$ 需要重新计算. 相应的推荐权重 $w_{i,j}$ 也根据新的相似程度和信任程度（新进入的用户信任程度初始化为 t_0）进行更新，然后返回步骤（1）.

这样，随着系统的运行和信任值的不断调整，推荐权重较高的用户逐渐稳定出现在顶部信任层 H^R 中，而与用户兴趣爱好不符的推荐者逐渐被置于缓冲层 H^B 中，并最终隔离在推荐产生的过程之外. 在大多数时间里，系统只需为每个用户维护 $2k$ 个用户集合. 并且随着系统的运行，无论是信任程度还是相似程度都根据用户的实际兴趣爱好得到了合理的调整. 推荐结果始终来自于最受其信任且相似度较高的用户群，从而有效的将攻击者隔离在推荐系统之外.

系统的流程如图 2 所示.

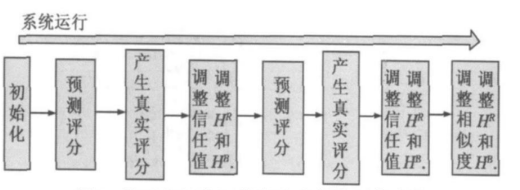


图2 基于信任分层的防攻击推荐系统流程

4 实验及结果分析

4.1 数据集

我们选取两个推荐算法研究中常见的数据集，分别为 Movielens 数据集* 与 Epinions 数据集**. 其中，Movielens 数据集包括 943 个用户在 1682 个电影上的 100，000 条评分记录；Epinions 数据集中包含 49，290 位用户在 139，738 个项目上的 644，824 条评分记录；二者的评分范围都是 1～5 分.

由于数据集是相当稀疏的，为了实验的顺利进行，我们进行了密集化处理. 对于两个数据集，取评分记录最多的 150 个用户，并抽取评分记录最多的 200 个项目，构成新的数据集. 处理后的 Movielens 数据集包含 19，064 条评分记录，

* http://www.grouplens.org/node/73

** http://www.trustlet.org/wiki/Epinions_datasets

$$\psi_M = 1 - \frac{19064}{150 \times 200} = 36.45\%;$$

处理后的 Epinions 数据集包含 6,621 条评分记录,

$$\psi_E = 1 - \frac{6621}{150 \times 200} = 77.93\%.$$

同时,我们将数据集划分为 5 份,采用 5 折交叉验证(5-fold cross-validation),即轮流将其中 4 份作为训练集,1 份作为测试集,取最终的平均值作为实验结果.

4.2　攻击数据

在下面的实验中,我们通过插入伪造的用户概要文件来攻击推荐算法.我们采用推动攻击和核攻击两种攻击策略,为了使攻击的效果更为明显,在推动攻击中,被推动的目标项目在伪造的概要文件被赋予最高的评价,而在核攻击中,被贬低的目标项目在伪造的概要文件被赋予最低的评价.攻击的概要文件的多少对推荐算法效果有重要的影响,实验中我们分别采用 5 ~ 40 条攻击数据,我们构造的攻击数据基于以下两点现实考虑:

(1)由于伪造用户概要文件严重影响电子商务系统推荐的效果,因此几乎所有的电子商务系统用户的注册都需要进行人力的介入(验证码),所有的评分都需要付出一定的代价(购买商品).

(2)现在的电子商务系统都有对伪造用户概要文件的检测机制,伪造概要文件的数量过多,特征过于明显都容易被系统发现并删除.

因此,真正具有攻击能力的攻击方法都倾向于需要较少的伪造用户概要文件.

4.3　评价标准

首先,我们使用平均绝对偏差 MAE(mean absolute error)来衡量推荐系统的预测精准度.假设系统对 N 个项目的预测评分向量为$\{p_1, p_2, \cdots, p_N\}$,用户实际评分为$\{r_1, r_2, \cdots, r_N\}$,则MAE定义为:

$$MAE = \frac{\sum_{i=1}^{N} |p_i - r_i|}{N}$$

MAE 值越小,代表推荐系统的精准度越高.

其次,我们使用平均预测偏离值 Δ(average prediction shift)来衡量推荐算法的鲁棒性.以 $p_{u,i}$ 表示用户 u 对项目 i 在没有推荐下的真实评分,$q_{u,i}$ 表示受到攻击影响后的评分,则平均预测偏离值 Δ 定义为:

$$\Delta = \frac{\sum_{u \in U} |q_{u,i} - p_{u,i}|}{|U|}$$

Δ 越接近于 0,说明推荐算法的鲁棒性越强.

最后,为了验证算法对恶意攻击的免疫力,设 A^R 为所有用户信任层 H^R 中的攻击者数目,A^B 为所有用户缓冲层 H^B 中的攻击者数目,我们定义最有可能在推荐中产生作用的攻击者平均信任值 $AvgTru_{\text{Att}}$ 为:

$$AugTru_{\text{Att}} = \frac{\sum tru_{u \to u_a}}{|A_R \cup A_B|}, \quad u \in |U|, |u_a| \in |A_R \cup A_B|$$

定义攻击者在信任层所占的比例 P_{Att} 为:$P_{\text{Att}} = \frac{|A_R|}{|H_R|}$.显然,这两个值越低,代表算法对恶意攻击的免疫力越强.

4.4　实验结果及分析

我们比较以下三种推荐算法的效果:

(1)传统的无信任关系的协作过滤推荐算法 CF;

(2)不分层的信任协作过滤推荐算法 TCF;

(3)基于信任分层的协作过滤推荐算法 TLCF;

对于后两个算法,初始信任值 $t_0 = 0.5$.

4.4.1　无攻击情况下的推荐质量比较

首先我们使用平均绝对偏差 MAE 来衡量推荐系统的预测精准度.图 3 给出了未受攻击的情况下三种算法的 MAE 值随参数 k 变化的趋势,其中 k 是系统用于产生推荐的最近邻个数(即信任层中的用户个数).在图 3 中,我们可以看到在 k 相同的情况下基于信任的协同推荐算法 TCF,TLCF 的准确性要好于无信任关系的协同推荐算法,因此信任度的引进对提高推荐性能有着重要的作用.而使用了分层的信任机制的 TLCF 算法推荐的准确性最高.

不分层的信任协作过滤推荐算法 TCF,只考虑调整目标用户 k 个最近邻的信任度.但是考虑到信任度的精确度对推荐结果的准确性起着至关重要的作用,TLCF 算法不但调整目标用户 k 个最近邻的信任度,而且调整目标用户缓冲层中的用户的信任值,保证推荐结果是由与目标用户权重最高的用户提供的.实验证明缓冲层的引入能够提高推荐的准确性.

由图 3 看出,$k > 35$ 后,MAE 值趋于平缓,为了权衡推荐质量和系统开销,在接下来的实验中,我们取 $k = 35$.

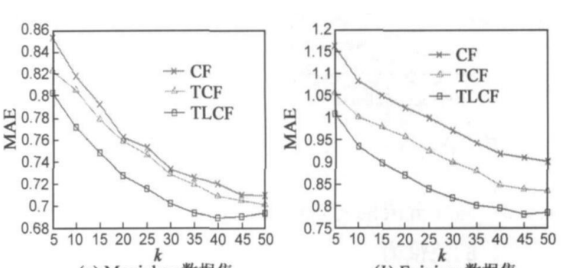


图3　三种算法在不同k值下的推荐效果

4.4.2　有攻击情况下推荐算法的鲁棒性比较

接下来我们通过平均预测偏移值 Δ 来考察有攻击情况下 3 个推荐算法的鲁棒性.Δ 越接近于 0,说明攻击对系统的影响越小.图 4 给出了在推动攻击中随攻击

数据的增加三个算法平均预测偏移值 Δ 的变化，在推动攻击中，平均预测偏移值 Δ 越大表明推荐算法受到攻击数据的影响越大，在图 4 中我们看到在两个数据集上 TLCF 算法的平均预测偏移值 Δ 在三个算法中都是最小的，当攻击数量达到 1/5 时，平均预测偏移值仍然小于 1. 图 5 给出了在核攻击中随攻击数据的增加三个算法平均预测偏移值 Δ 的变化，在核攻击中，平均预测偏移值 Δ 越小表明推荐算法收到攻击数据的影响越大. 在图 5 中，TLCF 算法依然表现出最好的性能，特别是在 Movielens 数据集中，核攻击对采用 TLCF 算法的推荐系统几乎没有产生什么影响.

通过实验我们看出在有攻击的情况下基于信任分层机制的 TLCF 算法的鲁棒性最好. 这是由于 TLCF 算法会根据推荐结果和用户的反馈调整目标用户信任层和缓冲层用户的信任度，将目标用户信任层中淘汰的用户放入目标用户的缓冲层，缓冲层中的攻击用户的信任度继续下降直到淘汰，而普通用户则有可能凭借准确的推荐再次进入信任层. 从而避免了将偏好与目标用户偏好不同的用户错误地识别为攻击用户，造成个性化信息丢失的后果，保证目标用户信任层的高权重.

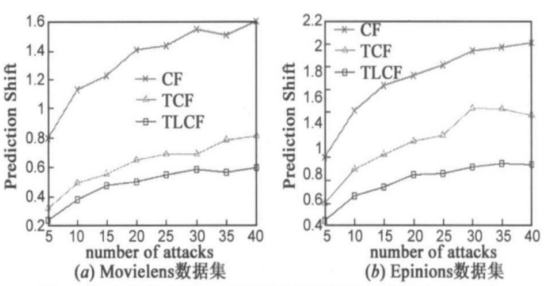


图4 推动攻击中攻击记录数与平均预测偏移值的关系

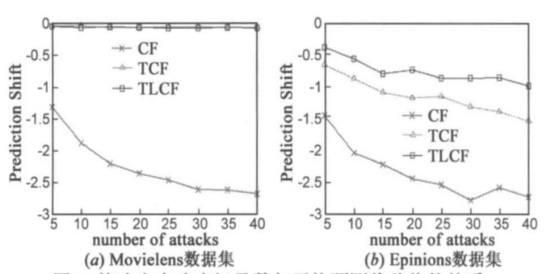


图5 核攻击中攻击记录数与平均预测偏移值的关系

4.4.3 信任分层的防攻击推荐算法 TLCF 对攻击数据的免疫力

最后，我们验证基于信任分层的防攻击推荐算法 TLCF 对攻击数据的免疫力. 我们将系统每次调整信任值及分层的过程称为一轮(round). 图 6 给出了在核攻击下(攻击记录数=40)随系统调整次数的增加，位于($H^R \cup H^B$)的攻击者平均信任值和位于 H^R 的攻击者比例的变化. 从图 6 中可以看出，随着算法运行时间的推进，每一轮 TLCF 算法都根据用户新的评分记录动态调整相应的信任值和信任分层，当算法运行 10 轮后，攻击者的平均信任度和位于 H^R 的攻击者的比例都显著下降，当算法运行 15 轮后攻击者的平均信任度和位于 H^R 的攻击者的比例都趋向于 0，从而失去攻击能力.

综上所述，在数据集 Movielens 与数据集 Epinions 上进行的实验都证明了引入信任机制确实能提高协作过滤推荐系统的推荐质量和防攻击能力，特别是采用了信任分层的 TLCF 算法，通过建立和维护目标用户的有限数量的信任用户和缓冲层用户，结合用户相似性来进行推荐，而且还利用用户新的评分记录对信任值和相似度进行动态调整，充分拟合了用户的兴趣爱好，能有效抵御推荐攻击的影响，并大大提高了推荐的质量.

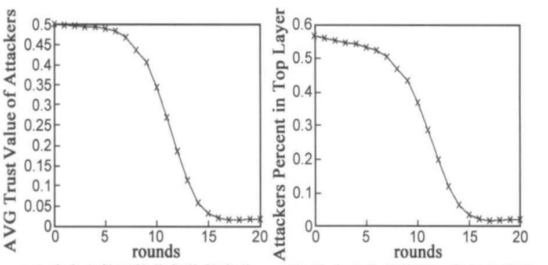


图6 核攻击中信任调整次数与攻击影响的关系

5 总结

本文介绍了协作过滤推荐系统攻击防御的研究背景与意义，阐述了协作过滤推荐与推荐攻击的基本原理. 在此基础上，结合社交网络中信任关系的思想，以用户对推荐结果的反馈为手段，动态调整用户间的信任关系，并将其通过分层的方式结合到协作过滤推荐系统中来. 实验证明，所述方法相对于传统的协作过滤推荐系统不仅提高了推荐的精准度，还大幅提高了推荐系统在攻击环境下的鲁棒性. 未来的工作包括考虑使用更复杂的模型处理包括信任关系、相似度关系、甚至是多种社会网络关系在内的用户间关系，进一步提高系统的精准度与鲁棒性；建立信任网络机制，以信任传递的方式缓解推荐系统冷启动问题.

参考文献

[1] 吴永辉，王晓龙，丁宇新，等. 基于主题的自适应、在线网络热点发现方法及新闻推荐系统[J]. 电子学报，2010，38(11)：2620—2624.
Wu Yong-hui, Wang Xiao-long, Ding Yu-xin, et al. Adaptive on-line web topic detection method for web news recommendation system[J]. Acta Electronica Sinica, 2010, 38(11): 2620—2624. (in Chinese)

[2] 张锋，孙雪冬，常会友，等. 两方参与的隐私保护协同过滤

推荐研究[J].电子学报,2009,37(1):84—89.
Zhang Feng, Sun Xue-dong, Chang Hui-you, et al. Research on privacy-preserving two-party collaborative filtering recommendation [J]. Acta Electronica Sinica, 2009, 37(1): 84—89. (in Chinese)

[3] 韩立新. 对搜索引擎中评分方法的研究[J].电子学报,2005,33(11):2094—2096.
Han Li-xin. A study on the ranking method of search Engines [J]. Acta Electronica Sinica, 2005, 33(11): 2094—2096. (in Chinese)

[4] Bamshad Mobasher, Robin D Burke, Runa Bhaumik, Chad Williams. Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness [J]. ACM Transactions on Internet Technology, 2007, 7(4): 23—38.

[5] P A Chirita, W Nejdl, C Zamfir. Preventing shilling attacks in online recommender systems[A]. Proceedings of ACM International Workshop on Web Information and Data Management [C]. New York, USA: ACM, 2005. 67—74.

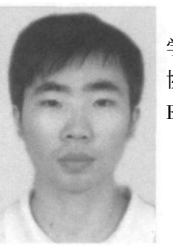
[6] Burke R, Mobasher B, Williams C, Bbaumik R. Segment-based injection attacks against collaborative filtering recommender systems[A]. Proceedings of the International Conference of Data Mining[C]. Chicago, USA: IEEE, 2005. 577—580.

[7] Nan Li, Chunping Li. Zero-Sum reward and punishment collaborative filtering recommendation algorithm[A]. Proceedings of the IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology[C]. Milan, Italy: IEEE, 2009. 548—551.

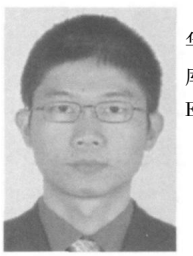
[8] H Ma, I King, M R Lyu. Learning to recommend with social trust ensemble[A]. Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval[C]. New York, USA: ACM, 2009. 203—210.

[9] H Yu, M Kaminsky, P B Gibbons, A Flaxman. SybilGuard: Defending against sybil attacks via social networks [J]. IEEE Transactions on Networking, 2008, 16(6): 576—589.

作者简介



黄世平　男,1987 年生于湖南邵东,中山大学信息科学与技术学院博士研究生.研究领域为协同软件技术、数据库、推荐系统.
E-mail: lship@mail2.sysu.edu.cn



黄　晋　男,1976 年生于海南琼中,博士,华南师范大学计算机学院讲师.研究领域为数据库理论、信息检索技术及推荐系统应用.
E-mail: dr.huangjin@gmail.com

陈　健(通讯作者)　女,1977 年生于广西柳州,博士,华南理工大学软件学院副教授、硕士导师.研究领域为数据库、数据挖掘、社会网络、推荐系统等.
E-mail: ellachen@scut.edu.cn