基于SVD的协同过滤算法的欺诈攻击行为分析

徐 翔,王煦法
XU Xiang,WANG Xu-fa

中国科学技术大学 计算机科学与技术系,合肥 230027
Department of Computer Science and Technology,University of Science and Technology of China,Hefei 230027,China
E-mail:xuustc@gmail.com

XU Xiang,WANG Xu-fa.Analysis of shilling attacks on SVD-based collaborative filtering algorithm.Computer Engineering and Applications,2009,45(20):92-95.

Abstract: Collaborative filtering is a vital central technology in personalized recommendation,but it is so sensitive to user profiles,that shilling attackers can easily inject biased profiles in an attempt to force a system to adapt in a manner advantageous to them.Recent research shows that the model and the cost of shilling attacks have different impacts on attack performance.This paper analyzes the attack effectiveness of different attack models on a SVD-based collaborative filtering algorithm,and the performances of attack models with different fill sizes and attack sizes using three evaluation parameters.

Key words: collaborative filtering;recommender systems;shilling attacks;Singular Value Decomposition(SVD)

摘 要:协同过滤是一种个性化推荐系统最常用的技术,但它对用户概貌信息较为敏感,欺诈攻击者很容易通过注入有偏差的用户概貌使系统的推荐结果有利于他们。研究表明欺诈攻击的攻击模型、攻击成本对攻击性能有不同程度的影响。针对这个问题,实验分析基于奇异值分解(SVD)的协同过滤算法在不同攻击模型下的性能表现,并以三种评估参数分析不同填充规模和攻击规模对攻击效率的影响。

关键词:协同过滤;推荐系统;欺诈攻击;奇异值分解
DOI:10.3778/j.issn.1002-8331.2009.20.028 **文章编号:**1002-8331(2009)20-0092-04 **文献标识码:**A **中图分类号:**TP393

1 引言

协同过滤算法是目前个性化推荐系统最常用的一种推荐算法,被广泛应用于电子商务。但电子商务竞争中经常出现一些不法用户为维护自身利益,向推荐系统中输入大量伪造评分数据,人为干预推荐系统的结果,导致系统的准确率下降。因此协同过滤系统的攻击行为研究日渐成为一个重要课题,是电子商务推荐系统安全应用的关键。

目前协同过滤算法主要分为三类:基于用户[1](user-based)的算法、基于项目[2](item-based)的算法和基于模型[3](model-based)的算法。国外学者对这三类算法的攻击研究已取得一定成果。Lam S[4]等人对user-based算法和item-based算法分别做了各种攻击测试分析,认为item-based算法比user-based算法有更好的抵御攻击的能力,以及推荐结果的表现方式对攻击效果有一定影响、并对新项目攻击问题提出一些建议。B Mobasher[4]则对基于模型的算法(K-means聚类和PLSA)和基于用户的算法(KNN算法)进行了攻击效果对比分析,认为PLSA和K-means聚类方法在系统稳定性和健壮性方面比user-based算法更胜一筹。本文在此基础上将研究另一种基于模型的协同过滤算法(SVD-based)对各种欺诈攻击的性能表现,以三种评估参数分析不同填充规模和攻击规模的攻击用户概貌对攻击效果的影响。

2 基于SVD的协同过滤算法

基于SVD的协同过滤算法由Sarwar[5]等人首次应用于协同过滤推荐中,他使用SVD方法将用户评分分解为不同的特征及这些特征对应的重要程度,利用用户与项目之间潜在的关系,用初始评分矩阵的奇异值分解去抽取一些本质的特征。SVD是矩阵维数简化的一种常用方法,它将一个 $m\times n$ 的矩阵 $\boldsymbol{R}$ 分解为3个矩阵。$\boldsymbol{R}=\boldsymbol{U}\times\boldsymbol{S}\times\boldsymbol{V}$,其中 $\boldsymbol{U}$ 是一个 $m\times m$ 的正交矩阵,$\boldsymbol{V}$ 是一个 $n\times n$ 的正交矩阵,$\boldsymbol{S}$ 是一个 $m\times n$ 的对角矩阵,它的对角线上的元素由上往下依次递减。

Sarwar通过将用户对未评分项的评分设为一个固定的缺省值来减少数据集的稀疏性。将矩阵 $\boldsymbol{R}$ 中评分值为0的项用相关列的项目评分平均值代替,接着将矩阵每行规范化为相同长度,用 $R_{ij}-\overline{R}_i$ 代替原来的 R_{ij}($\overline{R}_i$ 是第 i 个用户的平均评分值)。规范化为相同长度后,选择项目数较多的用户对相似度计算结果的影响降低了。经过这样的预处理得到矩阵 $\boldsymbol{R}'$,作为算法的输入矩阵。

作者简介:徐翔(1984-),男,硕士研究生,主要研究领域为网络信息处理,个性化推荐技术;王煦法(1948-),男,教授,主要研究领域为计算机网络、计算智能、信号处理和模式识别等。

收稿日期:2008-10-06 **修回日期:**2009-01-04

SVD 算法如下:

(1)用 SVD 方法分解矩阵 $\boldsymbol{R}'$得到矩阵 $\boldsymbol{U}$、$\boldsymbol{S}$、$\boldsymbol{V}$。

(2)将 $\boldsymbol{S}$ 简化为维数为 k 的矩阵,得到 $\boldsymbol{S}_k$($k<r$,r 是矩阵 $\boldsymbol{R}'$的秩)。

(3)相应简化矩阵 $\boldsymbol{U}$、$\boldsymbol{V}$ 得到 $\boldsymbol{U}_k$,$\boldsymbol{V}_k$。

(4)计算 $\boldsymbol{S}_k$ 的平方根得到 $\boldsymbol{S}_k^{1/2}$。

(5)计算两个相关矩阵 $\boldsymbol{U}_k\boldsymbol{S}_k^{1/2}$,$\boldsymbol{S}_k^{1/2}\boldsymbol{V}_k^{\mathrm{T}}$。

(6)每个用户 a 在未评分项目 i 上的预测评分为:

$$P_{a,i}=\bar{R}_a+\boldsymbol{U}_k\cdot\sqrt{\boldsymbol{S}_k}'(a)\cdot\sqrt{\boldsymbol{S}_k}\boldsymbol{V}_k'(i)$$

其中$\bar{R}_a$是用户 a 在所有已评分项目上评分的平均值。

3　欺诈攻击

3.1　定义

欺诈攻击指攻击者通过向推荐系统注入虚假用户,使系统的推荐结果产生偏差。每个攻击由多个虚假用户资料组成。一个攻击用户概貌(User Profile)[6]通常用一个 n 维向量表示,即 $\boldsymbol{UP}_i=(r_1,r_2,r_3,\cdots,r_n)$,其中 n 为系统内项目的总数量。记 I 为推荐系统的项目集,根据不同攻击模型的特点,I 由 I_T、I_S、I_F和 I_ϕ四部分组成,即 $I=I_T\cup I_S\cup I_F\cup I_\phi$。其中 I_T是攻击目标项目,I_S是为提高攻击效率而指定的项目集,对于某些攻击 I_S可以为空,I_F是需指定评分的项目集,I_ϕ是未评分的项目集。一次攻击由多个攻击用户资料组成,即 $Attack=(UP_1,UP_2,UP_3,\cdots,UP_{m-1},UP_m)$,其中 m 为攻击用户数量。

3.2　攻击意图

欺诈攻击按攻击意图划分可以分为两类,一类是 Push 攻击,另一类是 Nuke 攻击。前者的目的是提高目标项目被系统推荐的概率,后者刚好相反。电子商务推荐系统中常见的攻击类型为 Push 攻击,即恶意用户通过提交伪造的评分数据来提高目标项目的评分,以更高的概率推荐给顾客。

3.3　攻击成本

攻击成本反映攻击者实施攻击的难易程度及投入量,包括知识成本和执行成本[4]。知识成本指收集被攻击系统的信息及用户信息时所付出的努力;执行成本一般包括攻击规模(Attack Size)和填充规模(Fill Size),前者是攻击用户数量与系统中现有用户数量的比值,后者是每个攻击用户概貌中填充评分项数量与总项目数量的比值。

3.4　攻击模型

不同的攻击用户概貌导致攻击的效果不一,按攻击用户概貌的组成不同可以分为样本攻击、随机攻击、平均攻击、流行攻击等。这里主要分析随机攻击和平均攻击在基于 SVD 的协同过滤系统中的攻击效率。

3.4.1　随机攻击

随机攻击就是攻击者确定攻击目标后,选取一定填充规模的用户资料,使 I_F中所有项目的评价值在以所有用户对所有商品的平均评分值为中心的某个很小的范围内随机选取。很多系统中所有用户对所有商品的平均评价值是公开的,攻击者能够取得这些信息,因此随机攻击的知识成本是最小的,现实中是可行的。

3.4.2　平均攻击

平均攻击与随机攻击基本相同,I_F集中的项目 i 的评价值是在以所有用户对项目 i 的平均评价值为中心的某个很小的范围内随机选取的。它的知识成本较高,需要了解所有项目的评分平均值,实现难度较大。

3.5　攻击评估

评估各种攻击模式对系统的攻击效率一般有三种指标:MAE 偏差[4]、平均预测增量[6]和平均命中率偏差。

3.5.1　MAE 偏差

平均绝对偏差(MAE)是最常用的推荐质量度量方法,系统在受到攻击后,会对整个系统的推荐精度产生影响,MAE 也会出现偏差。因此在实验中采用攻击前后的 MAE 差值来衡量攻击效率,但在各种攻击下 MAE 变化很小,直观上很难表现不同类型攻击的效果差异。本文在各类攻击性能分析中仅将 MAE 偏差作为一个参考指标列出。

3.5.2　平均预测增量(Average Prediction Shift)

预测增量 $PS_{u,i}$描述用户 u 对项目 i 的预测评分在攻击后与攻击前的差值,一般对 Push 攻击其值为正,对 Nuke 攻击其值为负。为方便比较预测增量的数值变化,将其定义为正值,定义式为 $PS_{u,i}=\left|P'_{u,i}-P_{u,i}\right|$。令 U 和 I 为测试用户集和目标项目集,所有测试用户对项目 i 的平均预测增量为 $PS_i=\sum_{u\in U}PS_{u,i}/|U|$,类似的所有测试用户对所有目标项目的平均预测增量记为 $APS=\sum_{i\in I}PS_i/|I|$,一般 APS 越大,表明攻击越有效,反之表明系统越稳定。

3.5.3　平均命中率偏差(Average Hit Ratio Difference)

对 TopN 推荐形式的系统,采用平均命中率偏差来评估攻击效率。首先引入命中率[6]的概念,命中率是指测试集中目标项目 i 出现在 TopN 结果中的用户数与用户总数的百分比值。记 R_u表示系统向用户 u 推荐的 TopN 集合,$H_{ui}=1$ 表示项目 $i\in R_u$,否则 $H_{ui}=0$。测试集 U 对项目 i 的 Push 攻击的平均命中率定义为 $HitRatio_i=\sum_{u\in U}H_{ui}/|U|$。类似,测试用户集 U 对攻击项目集 I 的平均命中率为 $AHitRatio_i=\sum_{i\in I}HitRatio_i/|I|$。平均命中率偏差表示平均命中率在攻击后与攻击前的差异,记作 AHRD,与预测增量一样其值也有正负,为方便比较将其定义为正值。AHRD 越大,攻击越有效。

4　实验分析

实验将采用随机攻击和平均攻击两种模式对基于 SVD 的协同过滤系统进行攻击测试。实验选取了开放的公共数据集 MovieLens 的 10 万个评分数据作为算法的初始数据集,它由 943 个用户对 1 682 部电影的评分数据组成,实验随机选取其中的 459 个用户的 2 万个评分数据作为测试集 U。实验将分别采用不同的攻击类型和意图进行攻击测试,并对攻击效果进行评估分析。随机选取了不同评分数和平均评分的 20 个电影作为攻击目标项目集 T,所有实验均在 Matlab 平台上完成。首先介绍 SVD 算法中参数 k 的设定和攻击用户概貌的生成,然后用四类攻击(RandomPush、RandomNuke、AveragePush 和 Aver-

ageNuke)分别对系统进行攻击测试,并以 MAE 偏差、平均预测增量和平均命中率偏差为评估指标对攻击用户概貌的攻击规模和填充规模进行分析。

4.1 SVD 算法参数 k 的设定

SVD 算法中保留的维数 k 很重要,k 如果太小,不能得到用户评分矩阵的重要结构,k 太大则失去降维的意义,本节利用测试集数据先对 k 的取值进行优化,选取最优的 k 值然后再进行攻击实验。实验中 k 的取值从 1 到 25,步长为 1,分别计算 MAE,选取系统 MAE 最小时的 k 值作为最优值。k 与 MAE 的变化关系如图 1 所示,当 k=8 时最优,MAE=0.832 最小。

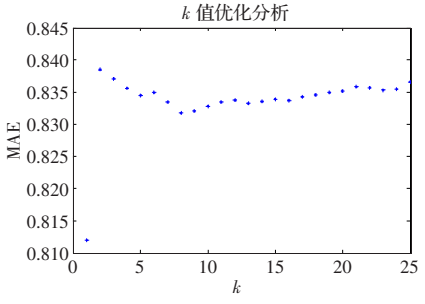

图 1 k 值优化分析

4.2 攻击用户概貌生成

对于随机攻击,测试集中的所有用户对所有项目的平均评分为 3.53,均方差为 1.1,故随机攻击的 UP 中 I_F的项目评分服从 N(3.53,1.1)的正态分布,I_T的评分由攻击意图而定,Push 攻击时 I_T=5,Nuke 攻击时 I_T=1,I_ϕ 中项目评分均为 0。对于平均攻击,先计算出测试集中的每个项目的平均评分 Avg_i,然后生成攻击者的用户概貌。UP 中 I_F的项目 i 的评分服从 $N(Avg_i,1.0)$ 的正态分布,其余设定同随机攻击。

4.3 三种评估参数对攻击规模和填充规模的分析

将以三种评估参数对 SVD 算法在不同攻击规模和不同填充规模情况下的攻击性能进行对比分析。实验将攻击规模和填充规模分别划分为 1%、5%、10%、15%、20%、30%、40%、50%、60%、70%、80%、90%、100%。实验以 RandomPush 攻击为例,总共对系统做了 169 次攻击实验,三种评估参数的实验结果分别见图 2、图 3 和图 4。

图 2 是平均预测增量与攻击规模及填充规模的关系变化图,在 FillSize 确定的情况下,平均预测增量随着 AttackSize 的增长幅度较快,增加近 1.3,可见平均预测增量受攻击规模的影响较大。但 AttackSize 在 30%以后平均预测增量缓慢增加,到 100%时只增加不到 0.2。当 AttackSize 确定时,FillSize 从 1%增加到 100%,平均预测增量只增加 0.05,几乎不受填充规模的影响。

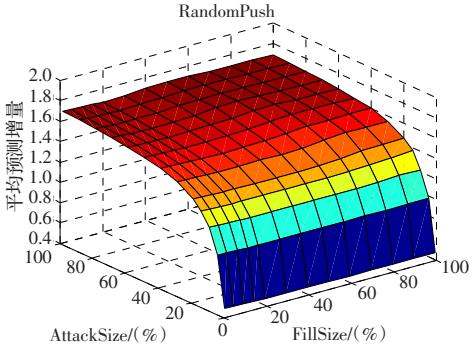

图 2 RandomPush 平均预测增量与填充规模及攻击规模的关系图

图 3 是 MAE 偏差与攻击规模和填充规模的关系变化图,如图所示,MAE 偏差在攻击规模和填充规模较小时不超过 0.08,只有当 AttackSize 大于 40%,填充规模大于 50%时,MAE 偏差在 0.12 附近。因此 MAE 偏差虽然不能精确体现系统受攻击的影响程度,但也一定程度上反映了攻击效果的变化趋势。

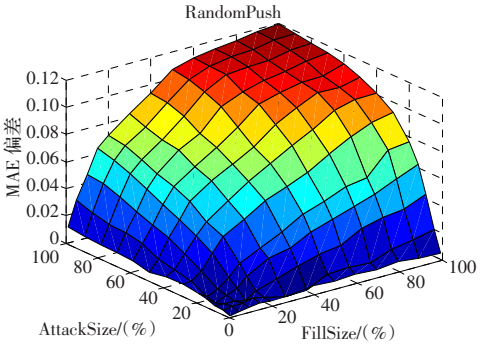

图 3 RandomPush MAE 偏差与填充规模及攻击规模的关系图

图 4 是平均命中率偏差与攻击规模和填充规模的关系变化图,系统的推荐形式为 Top50,由图可知当攻击规模在 1%~3%左右时,AHRD 受填充规模的影响在 2.8%到 5.5%之间徘徊,攻击对推荐结果影响效果较好,当攻击规模保持 5%以上时,AHRD 一直保持在 2.8%附近,几乎不受攻击规模和填充规模的影响。

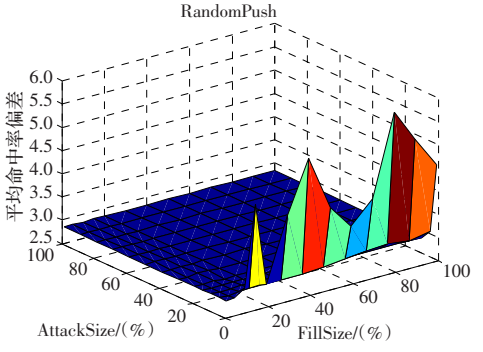

图 4 RandomPush 平均命中率偏差与填充规模及攻击规模的关系图

由上述结果可知,三种参数在衡量系统攻击效果的精度上有所差异。MAE 偏差对攻击的反应最为迟缓;APS 随攻击规模的增加而迅速增加,几乎不受填充规模的影响,但高的 APS 并不表明项目就会被推荐,还要看其他项目的评分;AHRD 反映了系统在攻击规模很小(1%~3%)时影响较大,攻击规模超过 3%后几乎不变。从评估的精度和直观性而言,AHRD 优于 APS,APS 优于 MAE 偏差。

实验还发现三种参数受攻击项目集选取差异的影响,如对于那些流行度较低的项目,攻击影响会比那些流行度高的项目更大。本文另选取了 20 个评分数在 5 以内的项目(平均评分在 0~2 之间)作为 RandomPush 项目集 PT,其实验结果与以 T 为项目集的结果对比,当攻击规模和填充规模均为 15%时 APS 提高近 1.7,AHRD 提高近 2%,MAE 偏差提升 0.015。由此可见 Push 攻击对那些评分较少的新项目攻击效果较好,这就迫切要求推荐系统解决新项目的安全性问题。

4.4　四种攻击类型对比分析

将分别在同等攻击规模不同填充规模的情况、同等填充规模不同攻击规模的情况下，讨论四种攻击类型对 SVD 算法的攻击性能差异,评估指标将采用 AHRD。

图 5 是 AttackSize=15%,填充规模变化的情况,RandomPush 和 AveragePush 的 AHRD 不受填充规模的影响稳定在 2.8%;RandomNuke 在较低填充规模(1%~5%)时 AHRD 较大,填充规模大于 10%时 AHRD 趋于稳定 2.8%,AverageNuke 在较低填充规模(1%)AHRD=19.5%和较高填充规模(>90%)时,AHRD=21%,在中等填充规模下 AHRD 变化范围为 15%~18%。

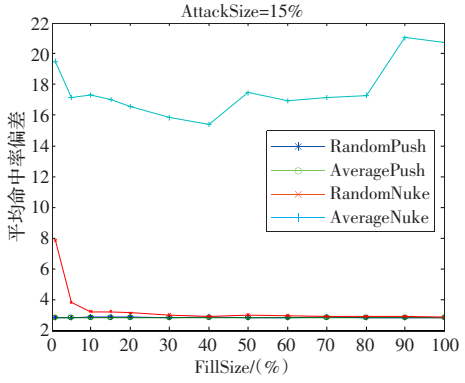

图 5　攻击规模为 15%时四种攻击类型的 AHRD 与填充规模的关系图

图 6 是 FillSize=15%,AttackSize 变化的情况,RandomPush、AveragePush 和 RandomNuke 的 AHRD 在较低攻击规模(1%~5%)时较高,但攻击效果相差不大;AverageNuke 在攻击规模=30%时 AHRD 达到最大值 24%,攻击效果最好。

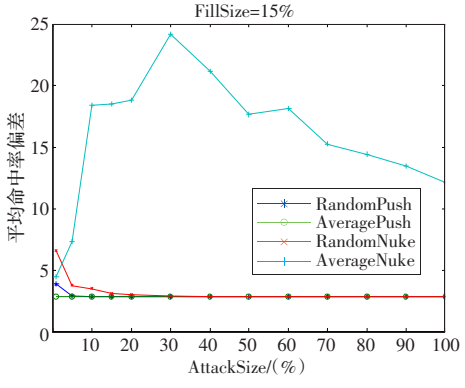

图 6　填充规模为 15%时四种攻击类型的 AHRD 与攻击规模的关系图

综上,对于两类 Push 攻击,在攻击规模小于 5%时,RandomPush 的攻击效果优于 AveragePush,但效果相差不大,攻击规模大于 5%,两类攻击效果几乎一致。对于两类 Nuke 攻击,当攻击规模小于 3%时,RandomNuke 的攻击效果优于 AverageNuke,但差距不大。当攻击规模大于 3%后 AverageNuke 明显优于 RandomNuke,且攻击效果高出许多。从整体而言,AverageNuke 的攻击效率远高于其他三类攻击，其 AHRD 比其他三类攻击高出近 15%;对于其他三类攻击,当攻击规模较低时(1%~5%),RandomNuke 的效果优于 RandomPush，优于 AveragePush,当攻击规模大于 20%时,这三类攻击效果非常接近,几乎不受攻击规模和填充规模的影响。

5　结语

针对一种基于 SVD 的协同过滤算法进行了四种攻击类型测试,采用三种评估参数对攻击用户概貌的填充规模和攻击规模对攻击效率的影响做了定量分析,同时对四种攻击类型的攻击效率做了对比分析。实验表明基于 SVD 的协同过滤系统对不同攻击规模和填充规模的 RandomPush、AveragePush 及 RandomNuke 攻击的防御能力较好，对 AverageNuke 攻击的防御能力较差。攻击样本的填充规模变化对 SVD 算法的性能影响微乎其微，原因可能是 SVD 对评分矩阵的初始化操作使得用户评分数目对用户相似度的影响降低。而攻击样本的攻击规模在较低的情况下能取得较好的攻击效果,较高的情况攻击效果变化不大。

参考文献:

[1] Resnick P,Iacovou N,Sushak M.GroupLens:An open architecture for collaborative filtering of netnews[C]//Proceedings of CSCW 1994, ACM SIG Computer Supported Cooperative Work,1994.

[2] Sarwar B,Karypis G,Konstan J,et al.Item-based collaborative filtering recommendation algorithms[C]//Proc of the 10th International WorldWideWeb Conference,2001:285-295.

[3] Mobasher B,Burke R,Sandvig J J.Model-based collaborative filtering as a defense against profile injection attacks[C]//Proceedings of the 21st National Conference on Artificial Intelligence(AAAI'06), 2006.

[4] Lam S,Reidl J.Shilling recommender systems for fun and profit[C]// Proceedings of the 13th International WWW Conference,New York, 2002.

[5] Sawrar B M,Karypis G,Konstan J A.Application of dimensionality reduction in recommender systems-A case study[C]//ACM WebKDD 2000 Web Mining for E-Commerce Workshop,Boston,Massachusetts,July 16-20,2006.

[6] Mobasher B,Burke R,Bhaumik R,et a.Effective attack models for shilling item-based collaborative filtering systems[C]//Proceedings of the 2005 Web KDD Workshop,Held in conjuction with ACM SIGKDD 2005,Chicago,Illinois,2005.

(上接 91 页)

[7] Zimmermann M G,Eguiluz V M.Cooperation,social networks,and the emergence of leadership in a prisoner's dilemma with adaptive local interactions[J].Phys Rev E,2005,72.

[8] Santos F C,Pacheco J M,Lenaerts T.Cooperation prevails when individuals adjust their social ties[J].PLOS Computational Biology, 2006,2:1284-1291.

[9] Pacheco J M,Traulsen A,Nowak M A.Active linking in evolutionary games[J].J Theor Biol,2006,243:437-443.

[10] Zachary W W.An information flow model for conflict and fission in small groups[J].Journal of Anthropological Research,1977,33: 452-473.

[11] Girvan M,Newman M E J.Community structure in social and biological networks[J].Proc Natl Acad Sci,2001,99(12):7821-7826.

[12] Newman M E J.Fast algorithm for detecting community structure in networks[J].Phys Rev E,2004,69(6).