·计算机工程与应用·

个性化推荐系统描述文件攻击检测方法

张 靖[1]，何发镁[2]，邱 云[3]

(1. 攀枝花学院网络中心 四川 攀枝花 617000; 2. 北京理工大学图书馆 北京 海淀区 100081;
3. 中国科学院成都计算机应用研究所 成都 610041)

【摘要】个性化推荐系统能产生针对性的、个性化的信息来满足不同用户需求，但也很容易受到用户描述文件注入恶意攻击，影响正常的推荐结果。针对该问题，分析和研究了描述文件的形式化模型、描述文件的属性及分类方法，应用粗糙集理论，设计了数据预处理离散化、决策表约简和个性化推荐处理相应算法，提出了一种用户描述文件分类学习和攻击检测的方法；为降低攻击对推荐结果的影响，完善了推荐系统的安全，设计出一种动态交互的个性化推荐模型框架。实例证明，用户描述文件的属性分类及检测方法是有效的，准确率高，能够有效地改善个性化推荐系统模型的安全。

关 键 词 分类; 描述文件; 检测; 推荐系统; 粗糙集理论
中图分类号 TP393.08 **文献标识码** A doi:10.3969/j.issn.1001-0548.2011.02.019

Inspection Method of the Attack on Personalized Recommendation System Description File

ZHANG Jing[1], HE Fa-mei[2], and QIU Yun[3]

(1. Computer Network Center, Panzhihua University Panzhihua Sichuan 617000;
2. Library, Beijing Institute of Technology Haidian Beijing 100081;
3. Chengdu Institute of Computer Application, Chinese Academy of Sciences Chengdu 610041)

Abstract Personalized recommendation system can satisfy the users' demand with pertinent and personalized information, but it is easy to be attacked maliciously by the description file, which will influence the recommendation result. The attribute, model, and classification method of the description files are analyzed and studied. The rough set theory is used to design an algorithm of data pretreatment discretization, decision table reduction, and personalized recommendation treatment. The method of description file classification and attack detection is proposed. The safety of the recommendation system is improved to decrease the influence of the attack on the recommendation results. The frame of personalized recommendation model with dynamic interaction is considered. The example verification proves that the model, the attribute classification, and the detection method of the description files are effective with high accuracy and can effectively improve the safety of the personalized recommendation system.

Key words classification; description file; inspection; recommender system; rough set theory

推荐系统是在与用户的交互过程中，使用动态的用户描述文件(即用户偏好等信息)产生个性化推荐结果。该系统作为开放式的应用系统，能够为用户和企业带来巨大的经济效益，但却受到恶意用户的关注。恶意用户期望通过影响推荐系统增加自己感兴趣项目的推荐频度，或降低竞争对手项目被推荐的可能性。近年来，一些研究者开始研究如何增强推荐系统的健壮性和稳定性，主要包括发现恶意攻击、采取有效的方法响应攻击以及信任推荐[1]等。描述文件注入攻击是推荐系统最常见的攻击，攻击者主要通过向推荐系统注入伪造的用户分级值，影响推荐系统的推荐结果或预测分级。为了确保协同过滤推荐系统产生有用的推荐或预测，必须及时发现描述文件注入攻击，并尽可能地降低其对协同过滤推荐系统性能的影响。

在研究识别可疑描述文件的过程中，可采用分

收稿日期：2010 – 05 – 17; 修回日期：2010 – 09 – 27
基金项目：四川省科技厅科研项目(2009zr0159)
作者简介：张 靖(1972 –)，男，副教授，主要从事软件工程和计算机网络方面的研究.

类模式匹配的方法检测攻击。文献[2]提出了用于分析恶意用户分级模式的指标，评价检测描述文件注入攻击的能力；文献[3]提出了一个扩散相似性算法，用来检测相似的攻击组；文献[4]通过实验量化地分析各种攻击对基于用户和基于项目的协同过滤推荐系统的性能影响；文献[5]开发了几种用于防范针对协同过滤推荐系统攻击的方法；文献[6]使用C4.5分类器对描述文件进行分类，基于C4.5决策树算法能够很好地发现攻击描述文件，但在大数据集的情况下，该方法不是很有效。为此，本文引入粗糙集理论，对推荐系统描述文件数据库进行分类，将攻击描述文件从推荐系统的描述文件数据库中分离出来。根据个性化推荐系统显著的动态性和交互性，针对描述文件注入攻击，分析研究用户描述文件的属性检测、分类、评价，提出了一种描述文件攻击检测方法，并在此基础上完善现有推荐系统模型，设计出一种动态交互安全推荐模型。

1　描述文件分类属性

1.1　描述文件形式化模型

定义 1　一个攻击模型是一个4元组 $M=\langle\chi,\delta,\sigma,\gamma\rangle$。其中，$\chi(i_t,I,U,\varphi)=\langle I_S,I_F,I_\varphi\rangle$ 是选择函数，包含一个目标项 i_t、所有项目的集合I、所有用户的集合U、参数集 φ；I_S 是一组被选择的项目，基于参数集 φ 中先前指定的参数决定；I_F 是一组随机选择的填充项目，基于参数集 φ 中预先指定的随机变量；$I_\varphi=I-(I_S\cup I_F\cup\{i_t\})$ 是一组未分级项目。$\delta:I\to R$ 和 $\sigma:I\to R$ 是项目集合I中元素的映射，分别用于给 I_S 和 I_F 赋予分级值；$\gamma:\{i_t\}\to R$ 是对目标项目 i_t 赋值的映射。

定义 2　一个基于攻击模型M的攻击描述文件是一组项目值对 $ap(M)=P_S\cup P_F\cup P_t\cup P_\varphi$。其中，$M=\langle\chi,\delta,\sigma,\gamma\rangle$ 是一个攻击模型；$P_S=\{\langle i,r\rangle\mid i\in I_S;\ r\in R,\delta(i)=r\}$；$P_F=\{\langle i,r\rangle\mid i\in I_F,\ r\in R,\sigma(i)=r\}$；$P_t=\{\langle i_t,r_t\rangle\}$，$r_t\in R$；$\gamma(i)=r_t$；$P_\varphi=\{\langle i,r\rangle\mid i\in I_\varphi,\ r=\text{null}\}$。

由定义1和定义2，描述文件注入攻击用户描述文件的形式如图1所示。

i_1^S	…	i_k^S	i_1^F	…	i_l^F	i_1^Φ	…	i_v^Φ	i_t
$\delta(i_1^S)$	…	$\delta(i_k^S)$	$\sigma(i_1^F)$	…	$\sigma(i_l^F)$	null	null	null	$\gamma(i_t)$
K个被选择项目的分级			l个填充项分级			攻击描述文件未分级项			目标项分级

图1　描述文件注入攻击模型的形式框架

1.2　描述文件属性

在分类中使用了通用属性、特定模型属性和内部描述文件属性。通用属性是根据基本描述性统计建立的，它企图获取一些倾向于是一个攻击者的描述文件看起来有别于正常用户的特征；特定模型的属性被设计来检测由特定攻击模型产生的描述文件的特征；内部描述文件属性被设计来检测描述文件之间的关系。

1) 通用属性[7]分别是：① 分级偏离平均度(RDMA)；② 权重度(WDA)；③ 权偏离平均度(WDMA)；④ 描述文件长度变化(LengthVar)；⑤ 邻居用户相似度(DegSim)；

2) 特定模型的属性[7]分别是：① 平均攻击检测模型–填充项平均变化(FMV)，填充项平均差异(FMD)是用户分级和平均分级之间差异的绝对值平均；② 段攻击检测模型–填充项平均目标差异(FMTD)；③ 关注目标模型(TMF)属性。

2　描述文件注入攻击检测方法

2.1　检测方法

首先根据各种反映描述文件特征的统计属性，建立相应的决策表，以不同的填充项大小和攻击大小、以及各种攻击模型，采用粗糙集方法发现攻击描述文件；其次修正现有的协同过滤推荐算法，清除攻击描述文件，降低攻击对推荐系统产生推荐质量的影响；最后，修正现有推荐模型，提出新的推荐系统模型。

根据描述文件的属性，使用以下15个属性。1) 6个通用属性：WDMA、RDMA、WDA、LengthVar、DegSim(k=450)、调整的DegSim′(k=2,d=963)；2) 6个平均攻击模型属性(3个为推攻击、3个为诋毁攻击)：FMV、FMD、PV(描述文件变化)；3) 2个段攻击模型属性(1个为push，1个为诋毁攻击)：FMTD；4) 1个目标检测模型属性：TMF。

2.2　用户描述文件分类

2.2.1　分类方法

在训练数据和测试数据中注入攻击数据，扩展后的训练数据通过粗糙集理论进行数据预处理、数据约简，提取可用于分类的决策规则，对注入攻击的测试数据进行划分，并对划分效果进行评估。使用粗糙集理论和技术进行分类学习将数据中对决策(即分类)影响不大的冗余属性或取值去掉，得到对决策起作用的核，从而得到约简后的决策规则(分类规则)。

2.2.2　数据预处理算法

信息熵离散化方法和Naieve离散化方法不需要额外的参数，直接根据信息表或数据库本身进行离

散，但是离散化后分类的效果没有使用Semi Naive Scaler算法的分类效果好，因此选择Semi Naive Scaler离散化算法，主要步骤如下。

对每个属性$a\in A$，令断点集合$C_a=\varnothing$，执行以下操作：

1) 对a的所有可能的取值集合V_a进行排序，使得相邻的对象(记录)x_i、x_{i+1}有$v_a^i\leqslant v_a^{i+1}$,v_a^i,$v_a^{i+1}\in V_a$；

2) 计算x_i、x_{i+1}所属的等价类对应的决策中出现频率最多的决策值的集合D_i和D_{i+1}：

$$D_i=\{v\in V_d\mid v=\arg\max_{v'}\mid\{x\in[x_i]_a\mid d(x)=v'\}\mid\}$$

$$D_{i+1}=\{v\in V_d\mid v=\arg\max_{v'}\mid\{x\in[x_j]_a\mid d(a)=v'\}\mid\}$$

3) if $((D_i\subseteq D_{i+1})\text{or}(D_{i+1}\subseteq D_i))$ ， then $C_a=C_a\bigcup\{(v_a^i+v_a^{i+1})/2\}$。

2.2.3 基于可辨识矩阵和遗传算法的约简算法

Skowron可辨识矩阵使决策表的核与约简的计算较简单。使用遗传算法进行论域空间有效搜索，减少分类器的数目[8-9]，算法主要步骤如下。

输入：$T=\{U, A, V, f\}$，精度控制值r，权重ρ，$0<r$，$\rho\leqslant 1$。

输出：分类规则。

$S=\{M_D(i,j)_{n\times n}\mid M_D(i,j)_{n\times n}\neq\varnothing\}$；//计算T的可辨识矩阵

P=InitializePopulation(T)；//初始化群体

continue = evaluate(P)；//评价群体是否满足要求

while (continue) //如果群体不满足要求，继续循环

Parents[1..3]←SelectParents(P)；//选择父母

Offspring[1]←Crossover(Parents[1])；//进行交叉

Offspring[2]←Mutation(Parents[2])；//进行变异

Offspring[3]←Inversion(Parents[3])；//进行染色体倒位

P=Recombine(P, Parents, Offspring)；//基因重组产生新的群体

continue=evaluate(P)；//评价群体是否满足要求

对决策表属性进行约简，就是要从原始的决策表条件属性中去掉不必要的属性，即冗余属性，从而提高系统的效率，并保证简化后的决策系统具有与原先系统一样的分类能力；统计重复的决策规则，删除重复的行，计算每条决策规则的核与约简。

2.2.4 分类准确性评价指标

分类准确性用于刻画全部数据的分类准确性，计算公式为：

$$\text{correctness}=|\text{positives}|/|U| \tag{1}$$

式中，|positives|表示正确分类样本数目，包括正确分类的正常描述文件和攻击描述文件数目；U表示论域；$|U|$表示全部样本或描述文件数目。

2.3 改进推荐算法

发现攻击之后，应将攻击从协同过滤推荐系统的描述文件数据库中移除，降低攻击对系统预测的影响。

使用传统Pearson 相似度计算方法：

$$\text{sim}_{u,v}=\frac{\sum_{i\in I}((r_{u,i}-\bar{r}_u)(r_{v,i}-\bar{r}_v))}{\sqrt{\sum_{i\in I}(r_{u,i}-\bar{r}_u)^2}\sqrt{\sum_{i\in I}(r_{v,i}-\bar{r})^2}} \tag{2}$$

式中，$\text{sim}_{u,v}$表示活动用户u和一个潜在邻居之间的相似性；$r_{u,i}$表示活动用户对一个项目的定级；$\bar{r}_u$表示用户u的所有定级的平均。

预期值为：

$$p_{u,i}=\bar{r}_u+\frac{\sum_{v\in V}\text{sim}_{u,v}\times(r_{v,i}-\bar{r}_v)}{\sum_{v\in V}|\text{sim}_{u,v}|} \tag{3}$$

考虑到描述文件数据库中存在攻击描述文件，引入一个变量PA表示一个描述文件被攻击的概率。当描述文件u被认为是正常描述文件时，$\text{PA}_u=0$，当某一描述文件u被认为是攻击时，$\text{PA}_u=1$，将其除去，降低它对推荐系统预测的影响，将相似度计算公式修改为：

$$\text{sim}'_{u,v}=\text{sim}_{u,v}\times(1-\text{PA}_v) \tag{4}$$

式中，PA_v表示描述文件v被攻击的概率。

相应地预测公式改进如下：

$$p'_{u,i}=\bar{r}_v+\frac{\sum_{v\in V}(\text{sim}'_{u,v}\times(r_{v,i}-\bar{r}_v))}{\sum_{v\in V}|\text{sim}'_{u,v}|} \tag{5}$$

由此，对协同过滤推荐算法进行改进[10]，主要步骤如下。

1) 对当前用户u要产生一个依照用户相似度大小进行排列的邻居的集合$N\{N_1, N_2,\cdots,N_m\}$，u不属于N，$\text{Sim}'(u, N_i)>\text{Sim}'(u, N_i+1)$。

2) 产生Top-N推荐集$V=\{N_1, N_2,\cdots, N_n\}$，满足条件$u$不属于$N$；$\text{Sim}'(u, N_i)>\text{Sim}'(u, N_i+1)$。

3) 在上述Top-N推荐集V的基础上，计算$p'_{u,i}$得到用户u对项目i的评级预测值。

2.4 实验及结果分析

实验采用MovieLens站点提供的100k评级数据(http://movielens.umn.edu/)，包括943个用户对1 682部电影的100 000个评分值。评级数据格式为user id|item id|rating|timestamp，分别为用户ID号、被评级电影的ID号、评级和时间戳。按照以下情况，实验分别验证对平均攻击(average attack)、随机攻击(random attack)、流行攻击(bandwagon attack)的检测

效果。1) 攻击大小(attack size)固定为1%的各种攻击模型检测研究，填充项大小(filler size)从3%到100%变化的情形; 2) 填充大小固定为3%，攻击大小从1%到10%变化的情形。

2.4.1　预处理数据

经过对原始实验数据进行统计整理，得到15个检测属性，加上用户标识(userID)和分类属性(class)构成一条关于某个用户评级数据的检测集。通过属性抽取，将原始的评级数据进行统计、汇总和转换，用粗糙集理论对数据进行预处理，用Semi Naive Scaler算法对测试和训练数据中的数值数据进行离散化，并保留数据属性原有的不可分辨关系。对训练数据经过数据预处理、数值约简、属性约简后得到约简集。在分类学习过程中产生的分类规则，可以使用分类准确性评价指标，对不会影响分类的规则过滤，规则数量极大降低，可提高学习效率。

2.4.2　攻击大小为1%时的实验

取1%的攻击量是为了在较小攻击量和较少攻击信息的情况下检测出攻击描述文件。在攻击大小为1%的情况下，注入各种填充大小(1%～100%)的攻击后，总的检测准确度(correctness)由于攻击的样本数量太少(1%)，以至于填充量大小的变化在3类攻击和填充量变化的18个实验中，总的检测准确度均为98.95%。

2.4.3　填充项大小为3%时的实验

由于填充大小固定在3%，提供的攻击信息较少。图1为总的分类准确率比较。由图1可见，在不同攻击大小的实验中，总的分类准确率达到了99%以上，说明利用粗糙集理论对描述文件进行分类判断是有效的。填充量固定的情况下，攻击量越大，获得的属性信息越多，也更容易将攻击检测出来。填充固定的情况下，攻击量较小时，平均攻击检测的准确度低于另外两种攻击的检测准确度，这可能是平均攻击在构造过程中填充项的取值分布造成的。在填充量大于6%时，3种攻击的检测准确度都增大到接近100%，该结果比文献[11]中相对应的实验结果(<90%)好。

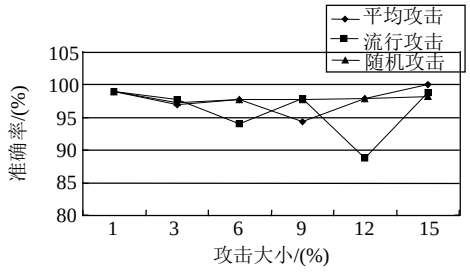

图1　总的分类准确率比较(filler size=3%)

图2为攻击描述文件的分类准确性比较。图中，攻击样本数量较少(attack size=1%，3%)时，获得的攻击信息较少，对平均攻击、随机攻击的检测准确度为0，即所有攻击都被误分成正常描述文件，但所有正常描述文件没有被误分为攻击。当攻击样本数量增加时，对分类的检测准确度呈跨越式上升(attack size=6%～15%)，当attack size=15%时，对平均攻击的检测准确率达到100%。流行攻击检测在填充量为6%、12%时，检测准确率均为0，这可能和填充数据的构造和填充项的分布有关系。

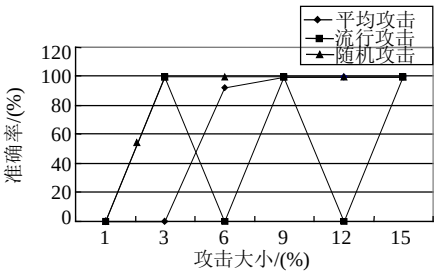

图2　攻击描述文件的分类准确性(filler size=3%)

图3为正常描述文件的分类准确性(filler size=30%)。从图3可见，填充量固定的情况下，攻击量越大，获得的属性信息越多，将攻击描述文件分类成正常描述文件的可能性也就越小。平均攻击在填充量为100%时，对分类标示为正常的描述文件的检测准确率为100%，没有一个攻击描述文件被误分为正常描述文件，该结果比文献[11]的实验结果(<99%)要好。随机攻击检测的准确度变化比较平稳，在98%附近有微小变化，结果与文献[11]相一致。在对流行攻击的检测实验中，把攻击记录分类成正常记录的几率比平均攻击和流行攻击高很多。

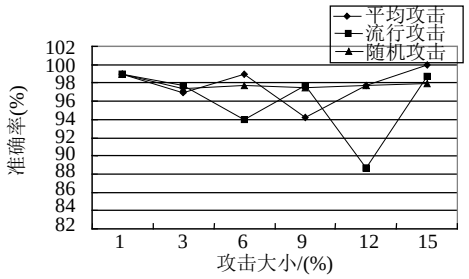

图3　正常描述文件的分类准确性(filler size=3%)

2.4.4　改进推荐算法后推荐影响实验结果

1) 最好情况：当filler size=3%，attack size=15%时，对Aver攻击进行检测，准确率达到100%，所有进行攻击的用户均被识别标示出来。对目标用户的推荐是基于其他真实用户的评级数据给出的，具有较好的可信性。

2) 最差情况：当attack size=3%时，对攻击的检测准确率为0，所有进行攻击用户均被标示为正常分

类，提交的评级数据将参与计算对目标用户推荐集的产生，这种分类学习结果对推荐系统有影响。正常数据产生推荐的平均绝对误差MAE=0.883 844，注入攻击后产生推荐的平均绝对误差 MAE′=0.887 045，ΔMAE=MAE – MAE′=0.003 201，可见，如果对数据库中注入恶意的评级数据，推荐性能仅仅下降了0.003 2，没有显著影响。

实验结果表明，对正常用户的评级准确率较高，可以利用粗糙集理论分类规则建立正常用户的使用模型。

3 完善推荐系统模型

使用分类的方法检测推荐系统描述文件数据库中的攻击描述文件，实验表明该方法能够有效地检测到攻击描述文件，可由此引入到推荐系统中进行攻击检测，完善推荐系统的安全。推荐系统具有动态性、交互性和整体性典型特征，根据P2DR模型思想，修正现有推荐模型，引入攻击检测安全机制，提出动态交互的安全推荐系统模型，提高了推荐系统的健壮性和稳定性。动态交互的安全推荐系统模型框架如图4所示。

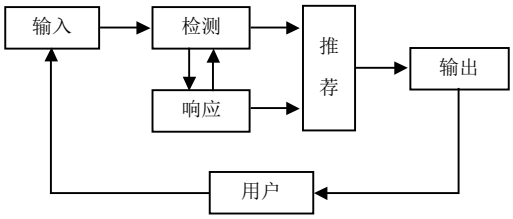

图4 动态交互的安全推荐系统模型框架

输入主要为推荐系统提供数据源。数据主要来源于用户信息，用户可以是客户个人和社团群体。输出主要把推荐系统产生的推荐结果输出给用户，有建议、预测、个体评分(评论)等。检测和响应是推荐系统安全的核心部分，在用户描述文件发生作用之前进行检测，如果有攻击描述文件被发现，则通过响应处理，清除该攻击文件。推荐是系统的主要部分，它使用经检测和响应处理过的用户描述文件，为用户产生预测或建议。

为了确保推荐系统的安全，要尽可能地发现攻击描述文件，并与推荐算法有效结合、与用户有效的信息交互，有效地将其从描述文件数据库中清除，最大程度地降低恶意用户攻击对推荐系统性能的影响。

4 结 论

通过分析、研究用户描述文件模型、属性和分类，结合粗糙集理论，对个性化推荐系统用户描述文件攻击提出了描述文件评级数据分类学习方法，为发现推荐系统中的攻击进行了分类学习。实验结果表明，评级数据分类方法在攻击大小合适的情况下(>6%)，能够较好地标示出攻击记录，尤其对正常数据分类的准确度几乎达到100%，为降低攻击对协同过滤推荐系统的影响，引入了攻击检测机制，完善了个性化推荐系统模型安全，具备了攻击检测响应机制，将攻击描述文件对系统预测的影响降到了最低程度。

参 考 文 献

[1] O'DONOVAN J, SMYTH B. Is trust robust?: an analysis of trust-based recommendation[C]//Proceedings of the 11th International Conference on Intelligent User Interfaces. New York, USA: ACM Press, 2006.

[2] CHIRITA P A, NEJDL W, ZAMFIR C. Preventing shilling attacks in online recommender systems[C]//Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management. New York, USA: ACM Press, 2005.

[3] SU Xue-feng, ZENG Hua-jun, CHEN Zheng. Finding group shilling in recommendation system[C]//Proceedings of the 14th International Conference on World Wide Web. New York, USA: ACM Press, 2005.

[4] BURKE R, MOBASHER B, WILLIAMS C, et al. Detecting profile injection attacks in collaborative recommender systems[C]//Proceedings of the 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE'06). Washington D C, USA: IEEE Computer Society, 2006.

[5] O'MAHONY M P, HURLEY N, KUSHMERICK N, et al. Collaborative recommendation: a robustness analysis[J]. ACM Transactions on Internet Technology, 2004, 4(4): 44-377.

[6] WITTEN I H, FRANK E. Data mining: practical machine learning tools and techniques[M]. 2nd ed. San Francisco, CA, USA: Morgan Kaufmann , 2005.

[7] BURKE R, MOBASHER B, WILLIAMS C, et al. Classification features for attack detection in collaborative recommender systems[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2006.

[8] 王国胤. Rough集理论与知识获取[M]. 西安: 西安交通大学出版社, 2001.
WANG Guo-yin. Rough set theory and knowledge acquisition [M]. Xi'an: Xi'an Jiaotong University Press, 2001.

[9] 邹瑞芝, 罗可, 曾正良. 基于粗糙集理论的决策树分类方法[J]. 计算机工程与科学, 2009, 31(10): 112-114.
ZOU Rui-zhi, LUO Ke, ZENG Zheng-liang. A classification method using decision trees based on rough sets[J]. Computer Engineering & Science, 2009, 31(10): 112-114.

[10] 何发镁, 王旭仁. 基于检测响应的安全协同推荐系统研究[J]. 微计算机信息, 2010, 26(6): 1-2.
HE Fa-mei, WANG Xu-ren. Research on secure collaborative recommender systems based on detection and response[J]. Microcomputer Information, 2010, 26(6): 1-2.

[11] WILLIAMS C. Profile injection attack detection for securing collaborative recommender systems[D]. Chicago: DePaul University, 2006.

编 辑 税 红