Journal of Computer Applications
计算机应用,2012,32(3):707 -709
ISSN 1001-9081
CODEN JYIIDU
2012-03-01
http://www.joca.cn

文章编号:1001 -9081(2012)03 -0707 -03
doi:10.3724/SP.J.1087.2012.00707

基于统计过程控制的协同推荐攻击检测方法

刘清林*,孟 珂,李苏丰
(中国矿业大学 计算机科学与技术学院,江苏 徐州 221116)
(*通信作者电子邮箱 tx048073@126.com)

摘 要:针对恶意攻击者利用协同推荐系统用户偏好敏感的缺陷向系统中注入虚假数据破坏推荐结果真实性的问题,提出基于统计过程控制(SPC)的协同推荐攻击检测方法。该方法将用户概貌项目评价数偏离度作为服务质量控制属性构建休哈特控制图,利用判异规则检测攻击用户,从而完善协同推荐系统模型。实验证明这种检测方法对各种不同的攻击模型都有较高的检测准确率和查全率。

关键词:协同推荐系统;统计过程控制;用户概貌项目评价数偏离度;托攻击;攻击检测
中图分类号:TP393.08 **文献标志码**:A

Attack detection method based on statistical process control in collaborative recommender system

LIU Qing-lin*, MENG Ke, LI Su-feng
(*School of Computer Science and Technology, China University of Mining and Technology, Xuzhou Jiangsu 221116, China*)

Abstract: Because of the open nature of collaborative recommender systems and their reliance on user-specified judgments for building profiles, an attacker could affect the prediction by injecting a lot of biased data. In order to keep the authenticity of recommendations, the attack detection method based on Statistical Process Control (SPC) was proposed. The method constructed the Shewhart control chart by using the users' deviation from the average of rating numbers and detected attackers according to the warning rules of the chart, thus improving the robustness of collaborative recommender systems. The experiments demonstrate that the method is effective with high precision and high recall against a variety of attack models.

Key words: collaborative recommender system; Statistical Process Control (SPC); deviation from average of rating number; shilling attack; attack detection

0 引言

推荐系统有效地解决了互联网发展带来的信息过载和信息迷航问题,特别是在电子商务领域,大部分的大型电子商务网站都在一定程度上使用了电子商务推荐系统。目前,主要的推荐系统有协同过滤推荐系统和基于内容的推荐系统两种[1]。由于基于内容的推荐系统无法解析信息的质量,难以区分资源内容的品质和风格,并且不能发现和更新与用户兴趣相似的资源等局限性和缺陷[2],对其的研究和应用都较少。基于协同过滤(Collaborative Filtering, CF)技术的推荐系统成为目前研究和应用最为广泛的个性化推荐技术。由于协同推荐系统对用户偏好信息的依赖,恶意攻击者可以通过向系统中注入虚假数据使系统频繁推荐其产品以谋求商业利益,使得推荐结果失去真实性,导致正常用户对推荐系统丧失信任和依赖,因此,保证推荐系统的安全性是非常重要的。

1 相关研究

协同推荐系统根据用户对项目的评分生成用户-项目评分矩阵,每个用户的评分行为称为用户概貌,协同推荐系统通过分析所有历史数据收集代表用户不同喜好的用户概貌,并根据与当前用户概貌最相近的邻居用户的评分行为产生推荐,这种推荐方式需要用户的直接参与,而且推荐结果依赖于用户偏好信息,这就给攻击者提供了可乘之机。攻击用户通过注入虚假用户评价信息,成为大部分用户的最近邻,从而试图改变推荐系统的推荐结果使其有利于自己的利益,这类攻击通常称之为用户概貌注入攻击(profile injection attack)或托攻击(shilling attack)[3]。目前研究较多的攻击模型有随机攻击(random attack)、平均攻击(average attack)、倾向攻击(bandwagon attack)等。

事实上,用户概貌注入攻击的产生也促使了推荐攻击检测模型的研究与发展。文献[4]提出了分析攻击用户评价行为的若干指标,用以检测某用户是攻击用户的潜在可能性;文献[5]提出在简化的攻击场景下,用一个扩散相似性算法来检测相似的攻击组;文献[6]从用户评价时间间隔入手,提出一种基于时间 SFM 因子的推荐攻击检测方法。另外还有一些针对特定攻击类型专有特性的攻击检测方法,如平均攻击检测模型、分类攻击检测模型等。但这些方法在大数据集、项目高填充量、适用范围、实际操作上都存在一定的局限性和不足。

协同推荐系统本身就是一个产生推荐的服务过程,攻击检测的目的就是通过隔离异常用户来保证推荐服务的质量。因此,本文从质量管理的角度出发,引入统计过程控制(Statistical Process Control, SPC)理论,从评价服务质量的角度入手,将用户概貌项目评价数偏离度作为服务质量好坏的

收稿日期:2011-09-07;**修回日期**:2011-11-14。

作者简介:刘清林(1985 -),男,山东泰安人,硕士研究生,主要研究方向:协同推荐系统; 孟珂(1987 -),男,江苏徐州人,硕士研究生,CCF 会员,主要研究方向:最短路径算法; 李苏丰(1982 -),男,江苏徐州人,硕士研究生,主要研究方向:电子商务、个性化服务。

评价指标，运用 SPC 理论中的控制图检测攻击用户，将符合判异准则的用户概貌视为攻击用户，由此提出一种协同推荐攻击检测方法，达到完善推荐系统模型的目的。通过针对最不易检测的随机攻击、平均攻击、倾向攻击 3 种攻击模型做大量实验来验证本文提出的攻击检测方法是可行且有效的。

2 基于 SPC 的协同推荐攻击检测方法

2.1 用户概貌项目评价数偏离度

统计过程控制是一种借助于数理统计方法的生产过程质量控制的重要工具，SPC 主要是指对生产、服务过程进行实时的质量监控，消除控制过程中产生的异常，恢复过程的稳定，从而使整个生产、服务过程处于可控状态。将 SPC 理论运用于协同推荐系统将攻击用户看作是服务过程异常，必须有一个能够区分正常用户与异常用户的质量特性值。

本文将用户概貌项目评价数偏离度（Deviation from Average of Rating Numbers）作为推荐服务过程的质量特性值。因为攻击用户必须针对目标项目注入大量伪造数据来达到成功攻击的目的，这就导致攻击用户概貌与真实用户概貌在数理统计特性上是有区别的，用户概貌项目评价数偏离度就是一个描述用户概貌差异的通用属性，本文使用用户概貌描述文件长度变化（*LengthVar*）表示。*LengthVar* 用于度量用户概貌项目评价数对整个数据库中用户平均评价数的偏离程度。真实用户概貌中的项目评价量不可能达到攻击用户概貌的项目评价量，使得 *LengthVar* 成为检测攻击用户概貌的最重要的指标[7]。*LengthVar* 的计算方法：

$$LengthVar_u = \frac{| n_u - \overline{n_u} |}{\sum_{u \in U} (n_u - \overline{n_u})^2}$$

其中：U 是所有用户集合，n_u 是用户 u 项目评价数，$\overline{n_u}$ 是整个推荐系统中用户项目评价数平均值。

2.2 协同推荐攻击检测中控制图的构建

SPC 一般由两个阶段组成。第一阶段分析所监控质量特性值的历史数据生成控制图，控制图要求所度量的质量特性值是可计量且基于随机分布的，由于每个用户之间评价行为是相互独立的，本文用 *LengthVar* 作为服务质量特性值，满足控制图的这一要求。根据常规控制图（休哈特控制图）[8] 的特点，在监控协同推荐系统时选用均值-极差控制图，图上有中心线（Center Line，CL）、上控制界限（Upper Control Limit，UCL）和下控制界限（Lower Control Limit，LCL），及质量特性值点列，如图 1 所示，将图中的质量特性值点与图中的上下控制限比较，就可以清晰地看出服务质量的变化。计算出 *UCL*、*CL*、*LCL* 就可以构建出控制图。

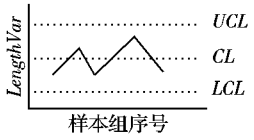

图 1 控制图的示意图

假设抽取 m 组用户样本，每组样本含 n 个用户概貌，每个用户概貌的 *LengthVar* 值作为其样本值，则控制图中 *UCL*、*CL*、*LCL* 的计算方法如下：

$$UCL_{\overline{X}} = \overline{X} + A_2 \overline{R}$$

$$CL_{\overline{X}} = \overline{X}$$

$$LCL_{\overline{X}} = \overline{X} - A_2 \overline{R}$$

其中：$A_2 = \frac{3}{d_2 \sqrt{n}}$，其值可以通过查计量控制图系数表得到；$\overline{X}$ 为样本中所有用户的总平均值；$\overline{R}$ 为样本中所有用户的极差平均值。$\overline{X}$ 和 $\overline{R}$ 的计算公式如下：

$$\overline{X} = \frac{1}{m} \sum_{i=1}^{m} \overline{x_i}$$

$$\overline{R} = \frac{1}{m} \sum_{i=1}^{m} R_i$$

其中：$\overline{x_i}$ 为本组样本中所有用户的平均 *LengthVar* 值，R_i 为每组样本的极差。

运用 SPC 的第二阶段，就是要利用休哈特判异规则[9] 来检测出攻击用户，判异规则为点出界就判异或界内点排列不随机判异。就协同推荐系统而言，当某组用户的 *LengthVar* 值落在控制线以外，则判定该用户为攻击用户，在产生推荐时就需要把该用户排除在外，不计算其与目标用户的相似性，从而保证系统推荐系统的服务质量。

2.3 完善的协同推荐系统模型

为完善协同推荐系统的安全性，本文提出在为用户推荐服务的过程中，使用 SPC 中的控制图监控整个服务过程，一旦发现攻击用户则立即将其去除，从而有效地减少攻击用户概貌在推荐产生过程中对推荐结果的影响。为达到剥离攻击者的目的，可以将相似性计算公式[10] 修改为：

$$sim'_{u,v} = sim_{u,v} * PA_v$$

其中：$sim_{u,v}$ 表示用户 u 和 v 的 Pearson 相似性；PA_v 表示用户 v 在向目标用户 u 产生推荐时是否有效，当监控到 v 为攻击用户时，$PA_v = 0$，攻击用户 v 不再在推荐中产生作用，从而达到剥离攻击者的目的。具有监控机制的协同推荐系统模型如图 2 所示。

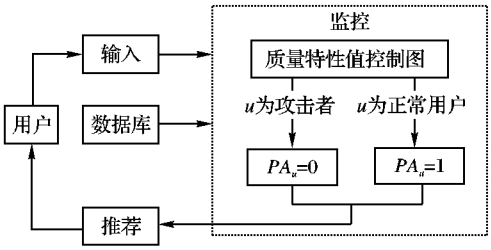

图 2 基于 SPC 的协同推荐系统模型

从图 2 可看出：监控部分是保证推荐系统安全的核心部分，该模型的最大好处是监控是自动完成的，用户在使用时根本觉察不到监控系统的存在，在保证安全推荐的前提下，不会给用户体验带来任何的负担。

3 实验与结果分析

3.1 评价指标

使用休哈特控制图监控服务过程，有可能出现两种错误：虚发警报和漏发警报。虚发警报是正常用户的 *LengthVar* 值超出界外，根据点出界判异而判为攻击用户；漏发警报是攻击用户的 *LengthVar* 值偶尔位于控制界限内而没有判异。为检验这两种错误的发生率，也是为了度量基于 SPC 的检测方法的有效性，本文采用准确率（*Precision*）和查全率（*Recall*）这两个指标来评价，计算公式如下：

$$Precision = N_t / N_a$$

$Recall = N_t/N_r$

其中:N_t 是检测到的真实攻击者的数量,N_a 是被检测方法标记为攻击者的数量(包括真实攻击者和误认为攻击者的正常用户),N_r 是整个系统中的真实攻击者数量。

3.2　实验设计与结果分析

本文使用开源的 MovieLens 100K 数据集(http://www.grouplens.org/node/12)作为实验数据集。此数据集包含 943 个用户对 1682 个电影的超过 100000 的评分值。所有的评分值都是介于 1~5 的整数值,其中 1 为最不喜欢,而 5 为最喜欢。数据集中所有用户均至少评价过 20 个以上的电影。

假设 MovieLens 数据集中不存在虚假用户评价行为,整个实验分为训练和测试两个阶段。在训练阶段,控制图判稳准则要求至少取 25 组数据,本文随机选取 30 组用户样本(多出 5 组以便发现不合适的数据可以舍去)作为训练集,每组样本含 5 个用户概貌,计算每组用户概貌的 *LengthVar* 平均值,做出 R 图。使用 R 图判稳后,构建的 $\bar{X}$ 控制图如图 3 所示。

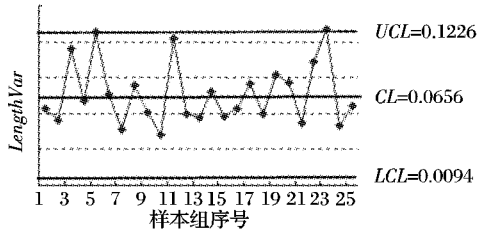

图 3　训练集 *LengthVar* 均值控制图

在真实的协同推荐系统中,托攻击的攻击大小达到 1% 已经是不可能的,本文将攻击大小提高到 3%,实验验证在填充规模从 10% 并以 10% 的幅度递增到 100% 的情形下对平均攻击、随机攻击、流行攻击的检测效果。图 4~5 分别描述了将攻击大小固定为 3% 时的检测准确率和查全率,从中可以看出:当填充规模小于 20% 时,随着填充规模的增加,准确率和查全率是一个下降趋势;当填充规模大于 20% 时,随着填充规模的增加,准确率和查全率也随之增加,但是即使在最低点查全率也能达到 98%,准确率则在 96% 以上;当填充规模低于 10% 或超过 90% 时,准确率和查全率都达到了 100%。同时,检测方法不受攻击模型的影响,对不同攻击模型的检测效果一直保持在稳定的状态,都有较高的准确率和查全率。

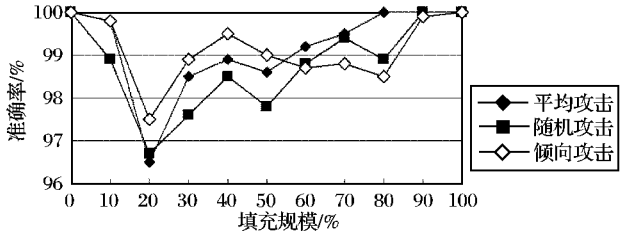

图 4　攻击大小为 3% 时的检测准确率

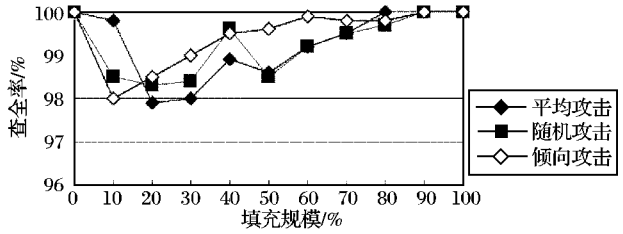

图 5　攻击大小为 3% 时的检测查全率

准确率和查全率的这种变化趋势是合理的,因为本文是将 *LengthVar* 作为质量控制特性,当攻击用户的填充规模与大部分用户项目评价数接近时,在均值控制图中,其点值就会落在控制中心线附近,从而导致检测方法误判。然而,攻击用户不可能知道全体用户的平均 *LengthVar*,即使碰巧攻击用户的填充规模达到了这个值,那么攻击就不会形成从而不会影响系统的推荐结果,因为攻击者需要较多评价项才能达到攻击效果,所以检测方法的这点不足是可以接受的,对检测效果影响是微乎其微的。更重要的是,该检测方法不受攻击模型种类的影响,主要原因是各种攻击模型都是根据用户对项目的评价值来区分的,而 *LengthVar* 与用户对项目的评价值没有关系。实验证明基于 SPC 的检测方法能大大保证推荐系统的服务质量,使其免受托攻击的困扰。

4　结语

统计过程控制的思想在工业产品质量控制中已经得到了广泛的应用,将其应用于协同推荐系统,通过监控用户概貌项目评价数偏离度及时发现并剔除攻击用户,保证了推荐系统的稳定性和推荐结果的真实性。实验证明基于统计过程控制的检测方法准确率和查全率高而且不受攻击模型种类的影响,能精确地检测出攻击用户,使向用户推荐的结果不受托攻击的影响,能充分保证推荐系统的服务质量。

参考文献:

[1] KWAK M, CHO D S. Collaborative filtering with automatic rating for recommendation[C]// ISIE 2001: IEEE International Symposium on Industrial Electronics. Piscataway, NJ: IEEE Press, 2001: 625-628.

[2] YU XIAOGAO, JIAN YIN. A new clustering algorithm based on KNN and DENCLUE[C]// Proceedings of ICMLC 2005. Piscataway, NJ: IEEE Press, 2005: 2033-2038.

[3] MOBASHER B, BURKE R, BHAUMIK R, *et al*. Effective attack models for shilling item-based collaborative filtering systems[C]// Proceedings of the WebKDD 2005. New York: ACM Press, 2005: 13-23.

[4] 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: ACM Press, 2005: 67-74.

[5] SU XUEFENG, ZENG HUAJUN, CHEN ZHENG. Finding group shilling in recommendation system[C]// Proceedings of the 14th International Conference on World Wide Web. New York: ACM Press, 2005: 960-961.

[6] 唐通. 基于时间 SFM 因子的推荐系统攻击检测方法[D]. 重庆:西南大学, 2010.

[7] WILLIAMS C, BHAUMIK R, BURKE R, *et al*. The impact of attack profile classification on the robustness of collaborative recommendation[C]// Proceedings of the WebKDD 2006. New York: ACM Press, 2006.

[8] 张公绪, 孙静. 统计过程控制与诊断[J]. 质量与可靠性, 2002(4): 39-45.

[9] SHEWART W A. Economic control of quality of manufactured product[M]. New York: Van Nostrand, 1931.

[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.