Neurocomputing 308 (2018) 205–226

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Uncovering anomalous rating behaviors for rating systems

Zhihai Yang*, Qindong Sun, Yaling Zhang, Beibei Zhang

School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China

ARTICLE INFO

Article history:
Received 18 December 2017
Revised 26 March 2018
Accepted 2 May 2018
Available online 8 May 2018

Communicated by Dr Xin Luo

Keywords:
Recommender system
Abnormality forensics
Shilling attack
Outlier detection

ABSTRACT

Personalization collaborative filtering recommendation plays a key component in online rating systems, which also suffers from profile injection attacks in reality. Although anomalous rating detection for online rating systems has attracted increasing attention in recent years, detection performance of the existing methods has not reached an end. Eliminating the impact of interfering information on anomaly detection is a crucial issue for reducing false alarm rates. Moreover, detecting anomalous ratings for unlabeled and real-world data is always a big challenge. In this paper, we investigate a two-stage detection framework to spot anomalous rating profiles. Firstly, interfering rating profiles are determined by comprehensively analyzing the distributions of user activity, item popularity and special ratings in order to eliminate sparse ratings. Based on the reserved rating profiles, combining target item analysis and non-linear structure clustering is then adopted to further determine the concerned attackers. Extensive experimental comparisons in diverse attacks demonstrate the effectiveness of the proposed method compared with competing benchmarks. Additionally, discovering interesting findings including anomalous ratings and items on two real-world datasets, Amazon and TripAdvisor, is also investigated.

1. Introduction

Online rating systems have been significantly developed in parallel with the social networks in the last decade. Rating data is ubiquitous on the well-known E-commerce websites including Amazon, Taobo, TripAdvisor, Yelp and etc. [5,6,11,12,21,32,48,55,61]. Personalization collaborative recommender systems play a crucial role in handling the increasingly prominent problem of information overload, which automatically suggest to a user items that might be of interest to her [1,4,23–28,41]. However, collaborative filtering recommender systems (CFRSs) are highly vulnerable to outside attacks, called profile injection attacks (a.k.a. shilling attacks) [6,34], due to the fact that recommender systems are entirely based on the input provided by users or customers [3,7,8,14,16,19,22,37,44,49,56,61]. Profile injection attacks, in which attackers manipulate biased ratings in order to influence future recommendations, have been demonstrated to be effective against collaborative filtering recommendation engines. According to the intention of attackers, shilling attacks can be classified in two basic categories: inserting malicious profiles which rate a particular item highly are called push attacks, conversely inserting malicious profiles aimed at downgrading the popularity of an item are termed nuke attacks [31]. Anonymous or pseudonymous users in online systems can multiply their profiles and identities nearly indefinitely, which utilize well-designed rating profiles to produce recommendation behaviors that the attackers desire. Therefore, proactively identifying the malicious rating profiles is extremely significant and meaningful for personalized collaborative recommendations.

Securing collaborative filtering recommender systems from malicious attacks have become an important issue with increasing popularity of recommender systems [13,47]. Although previous researches have shown promising results, defending such attacks is still an unresolved technique, and has not reached a full level of performance [13,36,47,57,61–64]. In particular, how to construct a strategy that can be used to spot anomalous ratings for real-world data is also extremely desired. Furthermore, developing detection method which can effectively defense diverse shilling attacks is always a big challenge. Moreover, compared with the number of genuine profiles (authentic profiles), the number of attack profiles is very small in rating systems. The distinct difference between the numbers of genuine and attack profiles is call imbalanced distribution of rating profiles [6,53]. The imbalanced distribution makes a challenging task for abnormality detection due to the difficulty of characterizing rating behaviors of users. How to eliminate a part of genuine profiles (interfering rating profiles) and reduce imbalanced distribution before anomaly detection is a concerned task especially for large-scale and real-world data. With respect to

* Corresponding author.
E-mail address: zhyang_xjtu@sina.com (Z. Yang).

https://doi.org/10.1016/j.neucom.2018.05.001
0925-2312/

unlabeled and real-world datasets, investigating abnormality forensics metrics for determining the concerned users or items is a realistic problem that cannot be ignored.

In this paper, we present a two-stage detection framework to spot anomalous rating profiles. Facing with the imbalanced distribution of rating profiles, interfering rating profiles are first determined by comprehensively analyzing the rating distribution of user activity, the distribution of item popularity and special ratings in order to eliminate sparse ratings. The goal of the first stage is to filter out interfering rating profiles (genuine profiles) [36,52] as many as possible and simultaneously reserve all attack profiles. Based on the remaining rating profiles, combining target item analysis and non-linear structure clustering is then adopted to further determine the concerned attackers. Since shilling attackers mimic rating details of authentic user to manipulate attack profiles, it is difficult to identify them. A robust multiple kernel data clustering method is employed to distinguish the attack profiles from authentic profiles in an appropriate feature space while the clusters are not linear separable in the original space. Moreover, we also explore evaluation metrics of abnormality forensics for discovering interesting findings in two real-world datasets including TripAdvisor and Amazon. More importantly, analyzing the internal relationship between historical ratings and reviews of items is provided to spot anomalous items. Extensive experimental comparisons on diverse attack datasets demonstrate the effectiveness of the proposed detection method compared with competing benchmarks. In addition, discovering interesting findings including anomalous items, ratings and etc. on Amazon and TripAdvisor datasets is investigated.

The main contributions of this paper are four-fold as follows:

- Eliminating interfering profiles according to the distributions of users' activity, items' popularity and sparse ratings in advance provides a feasible idea for abnormality detection faced with the imbalanced distribution of rating profiles, which is also favorable to characterize rating behaviors of users.
- Combining target item analysis and non-linear structure clustering is effective to reduce the scope of determining anomalous users. The false alarm rate of the proposed approach can be further reduced.
- To discover suspicious items or ratings on unlabeled and real-world datasets, suspected items detected by the proposed approach are further determined by comprehensively analyzing intrinsic association between overall rating and each aspect rating on the same item, rating behavior aggregation, rating intention distribution and topological structure analysis of suspicious items.
- Extensive experiments on both synthetic datasets in 10 different attacks and real-world datasets including Amazon and TripAdvisor are conducted to demonstrate the effectiveness of the proposed approach.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 introduces the proposed method in detail. In Section 4, experimental results are reported and analyzed. Finally, we briefly conclude the paper with a brief summary and discuss our future work.

2. Related work

Detecting anomalous rating behaviors has received much attentions over the last decade and achieved impressive results. In this section, we only discuss methods related to the presented work, which can be briefly introduced in the following three aspects, namely eliminating sparse ratings, clustering for shilling attack detection and anomalous rating detection for real-world data.

2.1. Eliminating sparse ratings

Anomaly detection can be considered as an unbalanced classification or an unbalanced clustering problem. Compared with the scale of genuine profiles, the number of attack profiles is relatively small in recommender systems. Based on the priori knowledge of attacks, eliminating interference rating profiles is favorable to reduce the detection range in advance. Morid and Shajari [36] developed a novel detection method in order to defense shilling attacks. They conjectured that the process of finding like-minded users forms a social network among all users and each link between two users represents an implicit connection between them. Users have more connections with others are the most influential users. The presented detection method only focuses on the influential users instead of the whole user set to improve their attack detection performance. Note that, the remaining users that are not influential users are considered as interference users, it is useful to narrow down the scope of detection. However, the precision and recall of the detection method are not impressive when the filler sizes are small. In addition, Zhou et al. [62–64] proposed stepwise detection methods to defense shilling attacks. They first analyzed rating patterns between malicious profiles and genuine profiles in attack model to preliminarily filter out interference profiles. Then, target item analysis effectively finds out suspected items, which is useful to further determine the concerned attack profiles. Recently, Yang et al. [51] presented a three-phase detection method to spot anomalous ratings. Note that, to eliminate interference ratings, constructing a user–user graph based on rating vector of user is useful to reduce the dimension of rating matrix in the first stage. However, only comparing the length of rating vector for each pair of users is time-consuming in reality especially for large-scale real-world datasets.

2.2. Clustering for shilling attack detection

Since shilling attackers mimic genuine users' rating details, the similarity between attackers is naturally higher than genuine users excluding the mimicked genuine users. In the previous researches, clustering-based detection frameworks are useful to capture the concerned attackers. Firstly, Mehta et al. [31,33] showed that clustering based on Principal Component Analysis (PCA) performed very well against standard attacks. The motivation behind this approach is that attacks consist of multiple profiles which are highly correlated with each other, as well as having high similarity with a large number of authentic profiles. However, while other attacks can be detected with high accuracy and fewer misclassified authentic users, performance of AOP attack detection [17] is not satisfactory. Bryan et al. [5] observed that the task of identifying attack profiles in recommender systems is similar to the task of identifying bi-clusters in gene microarray expression data. Besides, Bhaumik et al. [2] introduced an attribute-based k-means clustering approach to identify attack profiles regardless of attack types. The generic attributes used in the detection model which is based on descriptive statistics [35] are exploited to capture some of the characteristics that will tend to make an attacker's profile look different from a genuine user.

In addition, Lee and Zhu [20] proposed a hybrid two-phase method for detecting shilling attacks. A multidimensional scaling approach is first adopted to identify distinct behaviors that help to detect and secure the recommendation activities. Then, clustering-based methods are subsequently proposed to discriminate attack users. Cao et al. [7] proposed a MDS-based algorithm, which is a two-phase method. The detector first extracts a subset of effective users by computing a user–user dissimilarity matrix. Then, k-means is used to divide the selected users into K clusters. After that, Zhang et al. [59] proposed two clustering algorithms, CluTr