Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 12, December : 2023

Detecting Group Shilling Attacks in Online Recommender Systems Based on Bisecting K-Means Clustering

First Author: Sk Badulla, Assistant professor in Department of CSE, Ramireddy Subbarami Reddy (Rsr) Engineering College, Kadanuthala, Andhra Pradesh.

Second Author: V Sreenadh-, Assistant Professor in Department of CSE, Ramireddy Subbarami Reddy (Rsr) Engineering College, Kadanuthala, Andhra Pradesh

Abstract:

Recommender systems are highly vulnerable to shilling attacks, present shilling attack detection approaches focus mainly on identifying individual attackers in online recommender systems and rarely address the detection of group shilling attacks in which a group of attackers colludes to bias the output of an online recommender system by injecting fake profiles. In this article, we propose a group shilling attack detection method based on the bisecting K-means clustering algorithm. First, we extract the rating track of each item and divide the rating tracks to generate candidate groups according to a fixed time interval. Second, we propose item attention degree and user activity to calculate the suspicious degrees of candidate groups. Finally, we employ the bisecting K-means algorithm to cluster the candidate groups according to their suspicious degrees and obtain the attack groups.

1. Introduction

With the unstable development of online data, the marvel of data over-burden turns into a central point of contention. Online recommender frameworks make proposals for their clients, which can reduce the data over-burden issue partially. Be that as it may, the online recommender frameworks are defenseless against pushing assaults in which aggressors infuse countless assault profiles to predisposition the yield of the recommender framework. Peddling assaults can be separated into push assaults and nuke assaults, which are utilized for advancing and downgrading objective things (e.g., motion pictures or items) to be suggested, individually. The very much considered peddling assaults incorporate irregular assault, normal assault, fleeting trend assault, invert fad assault, normal objective shift assault, normal clamor infusing assault, etc. In these assaults, assailants ordinarily independently infuse assault profiles into recommender frameworks. Indeed, a gathering of aggressors may conspire to make a strategic assault. Such pushing practices have been named bunch peddling assaults, which are more threatening to the framework than customary pushing assaults. Hence, how to viably distinguish bunch pushing assaults has become a central question should have been tended to.

To secure recommender frameworks, different methodologies have been introduced to recognize peddling assaults over the previous decade. In any case, these methodologies center for the most part around distinguishing singular aggressors in recommender frameworks and infrequently consider the tricky pushing practices among assailants. Albeit a few methodologies have been proposed to distinguish peddling practices at the gathering level, they partition up-and-comer gatherings and recognize assault bunches as per profile likeness. There are some gathering assault models that can create assault profiles with incredible variety. Thus, these methodologies can't completely identify assault gatherings, which cause helpless exactness and review. As of late, a few methodologies have been introduced to identify spammer bunches in audit sites. Nonetheless, the gathering peddling assaults in recommender frameworks are not quite the same as the spammer bunches in audit sites. Subsequently, the spammer bunch location approaches are not pertinent to the identification of gathering peddling assaults

2. Literature survey

The detection of shilling attacks has been studied extensively over the past decade. The detection methods for shilling attacks can be divided into two categories: supervised methods and unsupervised methods. Supervised methods (e.g., kNN-C4.5-, and SVM-based detection algorithms) first use a large number of labeled instances to train a classification model, and then, the model is used for classifying attack profiles. Zhou *et al.* presented a two-step detection method based on SVM. They first utilized Borderline-SMOTE to relieve the unbalance classification situation and obtained a preliminary result via SVM. Then, they used a target item analysis method to identify attackers. Li *et al.* extracted some features from the item popularity degree and detected shilling attacks using the improved ID3 decision tree. This approach is not very effective when the filler size and attack size are small. The abovementioned approaches need to label sample data and train classification model, which are only applicable to detecting known types of shilling attacks. To overcome the limitations of supervised methods, some unsupervised methods have been proposed. Mehta and Nejdl [15] analyzed the similarity structure in attack profiles and used principal component analysis (PCA) to identify the attack profiles. Bryan *et al.* [16] utilized H-score to sort users and obtained the target items on the basis of the sorted list of users. After the two steps, attack profiles were detected by the target item deviation. Unsupervised detection methods do not need to consider attack types or label training samples, but they need *a priori* knowledge of attacks (e.g., the attack size).

The aforementioned methods focus mainly on detecting individual attackers in recommender systems. However, a group of attackers might collude to bias the output of recommender systems. Therefore, the detection of group shilling attacks has attracted attention in recent years. Zhou et al. improved the individual shilling attack detection metrics and proposed a two-step method to detect group shilling attacks. While this method is effective for detecting group attacks in synthetic data sets, it is not effective in detecting group attacks with a low similarity between attackers. Wang et al. improved several traditional features and proposed a method for group attack detection based on these features. They first manually labeled candidate groups with high minimum support, and thereafter, they computed the group metrics and employed PCA to rank the candidate groups. Unfortunately, this method is only suitable for detecting group attacks whose shilling profiles have high similarity.

Another work related to group shilling attack detection is the detection of spammer groups in review websites. Existing methods for detecting spammer groups can be divided into group content and behavior analysis-based approaches and group structure analysis-based approaches. For group content and behavior analysis-based approaches, features of group review contents and user behaviors are extracted for detecting spammer groups. For group structure analysis-based approaches, the network structure characteristics of groups are used for spammer group detection. As the group shilling attacks in recommender systems are more complex than the spammer groups in review websites, the spammer group detection approaches are not applicable to detecting group shilling attacks.

3. Proposed Work

We propose a method to detect group shilling attacks in online recommender systems through bisecting K-means clustering. The proposed approach takes advantage of the time concentration characteristics of group shilling attacks, which has a better performance in detecting group attacks with collusive shilling behaviors.

We propose a candidate group division method, which first mines the rating tracks of items and then divides the users in the item rating tracks (IRTs) into multiple groups according to a certain length of time. Since the attackers in an attack group must rate the target item(s) within a certain period of time, the proposed candidate group division method is more likely to divide the attackers in an attack group together, which can lay a good foundation for the group shilling attack detection.

We propose metrics of item attention degree and user activity (UA) to analyze the candidate groups, making the judgment of attack groups more accurate. Based on the divided candidate groups, the item attention degree and the UA for each candidate group are calculated, and the suspicious degrees of these groups are obtained. Based on this, the bisecting K-means algorithm is employed to cluster the candidate groups according to their suspicious degrees, and the attack groups are obtained.

To evaluate the performance of our method, we conduct experiments on the Netflix and Amazon data sets and compare the proposed method with four baseline methods.

Group Shilling Attacks

The concept of group shilling attacks was proposed by Su et al.. They provided two scenarios for such attacks. In scenario 1, besides giving biased ratings for the target item(s), the attackers also provide normal ratings for non target items to conceal their attack intentions. In scenario 2, the gray organizations first collect different target items and send these items to the hired members, and thereafter, the group members select some target items for attacking.

Bisecting K-Means Clustering Algorithm

The core idea of the bisecting K-means clustering algorithm is to treat all data samples as a cluster at first and then divide this cluster into two clusters. Subsequently, the cluster that can minimize the clustering cost function (i.e., the sum of squared errors) is selected and divided into two clusters. This process continues until the number of clusters reaches the given number K .

1) Use the basic K-means algorithm to divide all data samples into two clusters and add them to the set of clusters.

2) In the set of clusters, select the cluster that can reduce the sum of squared errors to the greatest extent and use the basic K-means clustering algorithm to divide it into two clusters, and, thereafter, add them to the set of clusters.

3) Repeat step 2) until there are K clusters in the set.

Unlike the traditional K-means clustering, the bisecting K-means clustering can overcome the situation that the algorithm enters the local optimal state to some extent.

Working Model

When a group of attackers colludes to mount an attack against the recommender system, they not only rate the target item(s) but also rate some non target items as well. Moreover, attackers in the group should complete their rating tasks within a certain period of time in order to achieve the desired attack effect. Based on such considerations, we propose a group shilling attack detection method based on bisecting K-means clustering, which is called GD-BKM.

The first step is to obtain candidate groups, and the users who rate the same item within a time interval are divided into the same group. In the second step, user features and item features are extracted. These features are combined to compute the suspicious degree of each candidate group. The last step is to use the bisecting K-means algorithm to distinguish attack groups according to their suspicious degrees.

Dividing Candidate Groups

In this section, each item’s rating track is constructed. Based on the rating tracks and a given time interval length (TIL) threshold, the candidate groups are generated.

Calculating the Suspicious Degree of Each Candidate Group

From the item perspective, the intent of an attack group is to increase the recommended probability of the target item. If attackers collude to promote or demote an item, the item’s attention degree will be high. To achieve the desired attack effect, attackers in an attack group are required to complete their rating tasks within a specified time, so the attackers in the group will be active in this time interval.

Detecting Attack Groups

Based on the divided candidate groups, we employ the bisecting K-means algorithm to cluster the candidate groups according to their suspicious degrees and identify the attack groups from the generated clusters of candidate groups.

4. Conclusion

Group shilling attacks are a great threat to recommender systems. To detect such attacks, we propose a group attack detection model based on the bisecting K-means algorithm. The proposed detection model can overcome the problem that the performance is poor when attackers have a few coated items. In order to divide candidate groups, we use the fixed time length and dynamically select the starting time point to divide each item's rating track. the bisecting K-means algorithm is utilized to identify attack groups from the candidate groups.

5. References

[1] T. L. Ngo-Ye and A. P. Sinha, "Analyzing online review helpfulness using a regressional relief F-Enhanced text mining method," ACM Trans. Manage. Inf. Syst., vol. 3, no. 2, pp. 10:1–10:20, Jul. 2012.

[2] D. Jia, C. Zeng, Z. Y. Peng, P. Cheng, Z. M. Yang, and Z. Lu, "A user preference based automatic potential group generation method for social media sharing and recommendation," (in Chinese) Jisuanji Xuebao, vol. 35, no. 11, pp. 2382–2391, Nov. 2012.

[3] I. Gunes, C. Kaleli, A. Bilge, and H. Polat, "Shilling attacks against recommender systems: A comprehensive survey," Artif. Intell. Rev., vol. 42, no. 4, pp. 767–799, Dec. 2014.

[4] S. K. Lam and J. Riedl, "Shilling recommender systems for fun and profit," in Proc. 13th Conf. World Wide Web WWW, 2004, pp. 393–402.

[5] B. Mobasher, R. Burke, R. Bhaumik, and J. J. Sandvig, "Attacks and remedies in collaborative recommendation," IEEE Intell. Syst., vol. 22, no. 3, pp. 56–63, May 2007.

[6] B. Mobasher, R. Burke, R. Bhaumik, and C. Williams, "Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness," ACM Trans. Internet Technol., vol. 7, no. 4, p. 23, Oct. 2007.

[7] C. Williams, B. Mobasher, R. Burke, J. Sandvig, and R. Bhaumik, "Detection of obfuscated attacks in collaborative recommender systems," in Proc. 17th Eur. Conf. Artif. Intell., 2006, pp. 19–23.

[8] X.-F. Su, H.-J. Zeng, and Z. Chen, "Finding group shilling in recom-mendation system," in Proc. Special Interest Tracks Posters 14th Int. Conf. World Wide Web WWW, 2005, pp. 960–961.

[9] R. Burke, B. Mobasher, C. Williams, and R. Bhaumik, "Classification features for attack detection in collaborative recommender systems," in Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD, 2006, pp. 542–547.

[10] Y. Wang, Z. Wu, J. Cao, and C. Fang, "Towards a tricksy group shilling attack model against recommender systems," in Proc. 8th Int. Conf. Adv. Data Min. Appl., Nanjing, China, 2012, pp. 675–688.

[11] K. Murugesan and J. Zhang, "Hybrid bisect K-Means clustering algorithm," in Proc. Int. Conf. Bus. Comput. Global Informatization, Jul. 2011, pp. 216–219.

[12] C. A. Williams, B. Mobasher, and R. Burke, "Defending recommender systems: Detection of profile injection attacks," Service Oriented Com-put. Appl., vol. 1, no. 3, pp. 157–170, Oct. 2007.

[13] W. Zhou, J. Wen, Q. Xiong, M. Gao, and J. Zeng, "SVM-TIA a shilling attack detection method based on SVM and target item analysis in recommender systems," Neurocomputing, vol. 210, pp. 197–205, Oct. 2016.

[14] W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and B. Ling, "An shilling attack detection algorithm based on popularity degree features," (in Chinese) Acta Automatica Sinica, vol. 41, no. 9, pp. 1563–1575, Sep. 2015.

[15] B. Mehta and W. Nejdl, "Unsupervised strategies for shilling detection and robust collaborative filtering," User Model. User-Adapted Interact., vol. 19, nos. 1–2, pp. 65–97, Feb. 2009.

[16] K. Bryan, M. O'Mahony, and P. Cunningham, "Unsupervised retrieval of attack profiles in

collaborative recommender systems," in Proc. ACM Conf. Recommender Syst. RecSys, 2008, pp. 155–162.

[17] W. Zhou, Y. S. Koh, J. Wen, S. Alam, and G. Dobbie, "Detection of abnormal profiles on group attacks in recommender systems," in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. SIGIR, 2014,pp. 955–958.