Journal of Engineering Sciences
Vol 12, Issue 05, MAY /2021
ISSN NO:0377-9254

Detecting Group Shilling Attacks In Online Recommender Systems

[1] B.Sharmila,[2] D.Narmada, [3]J. Krishna Keerthana, [4] K.L.Mounika

B.Sharmila Assistant Professor in Department of CSE Vignan institute of technology and science,Hyderabad,TS.

D.Narmada B.Tech Student in Department of CSE Vignan institute of technology and science,Hyderabad,TS.

J. Krishna Keerthana B.Tech Student in Department of CSE Vignan institute of technology and science,Hyderabad,TS.

K.L.Mounika B.Tech Student in Department of CSE Vignan institute of technology and science,Hyderabad,TS.

Abstract:

Existing 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 dB scan 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 dB scan algorithm to cluster the candidate groups according to their suspicious degrees and obtain the attack groups. The results of experiments on the Netflix and Amazon data sets indicate that the proposed method outperforms the baseline methods.

I. Introduction

With the explosive growth of online information, the phenomenon of information overload becomes a key issue. Online recommender systems make recommendations for their users, which can alleviate the information overload problem to some extent. However, the online recommender systems are vulnerable to shilling attacks in which attackers inject a large number of attack profiles to bias the output of the recommender system, . Shilling attacks can be divided into push attacks and nuke attacks, which are used for promoting and demoting target items (e.g., movies or products) to be recommended, respectively. The well-studied shilling attacks include random attack, average attack, bandwagon attack, reverse bandwagon attack, average-target shift attack, average-noise injecting attack, and so on. In these attacks, attackers usually separately inject attack profiles into recommender systems. In fact, a group of attackers might collude to make a tactical attack. Such shilling behaviors have

been termed group shilling attacks, which are more threatening to the system than traditional shilling attacks. Therefore, how to effectively identify group shilling attacks has become a key issue needed to be addressed To protect recommender systems, various approaches have been presented to detect shilling attacks over the past decade. However, these approaches focus mainly on detecting individual attackers in recommender systems and rarely consider the collusive shilling behaviors among attackers. Although some approaches have been proposed to detect shilling behaviors at the group level, they divide candidate groups and identify attack groups according to pro□le similarity. There are some group attack models that can generate attack pro□les with great diversity. As a result, these approaches canno t fully detect attack groups, which causes poor precision and recall. Recently, some approaches have been presented to detect spammer groups in review websites. However, the group shilling attacks in recommender systems are different from the spammer groups in review websites. Therefore, the spammer group detection approaches are not applicable to the detection of group shilling attacks. To overcome the abovementioned limitations, we propose a method to detect group shilling attacks in online recommender systems through dB scan algorithm. The proposed approach takes advantage of he time concentration characteristics of group shilling attacks, which has a better performance in detecting group attacks with collusive shilling behaviors. The major contributions of this article are listed as follows.

1. We propose a candidate group division method, which □rst 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.

2. 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, dB scan algorithm is employed to cluster the candidate groups according to their suspicious degrees, and the attack groups are obtained.

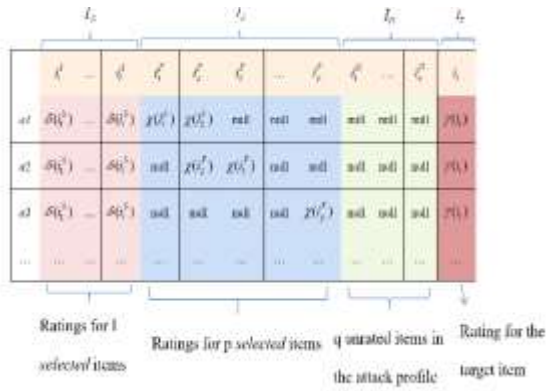
Objective:

Group shilling attacks are a great threat to recommender systems. To detect such attacks, we propose a group attack detection model based on the dB scan algorithm. The proposed detection model can overcome the problem that the performance is poor when attackers have a few corated 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.

II. RELATED WORK

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 nontarget 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. To achieve the desired effect of the attack, a fair amount of attack profiles must be injected into the recommender system. Fig. 1 shows one form of group attack profiles. In Fig. 1, IS is the set of selected items. Function δ determines the ratings of these items. There may be no selected items in some group attack profiles. IF is the set of filler items whose ratings provided by attacker i in the attack group. Similarly, Function χi decides ratings of filler items. In group shilling attacks, the filler items are very important. The filler items in the traditional shilling attacks are chosen randomly. In contrast, the selection of filler items in the group shilling attacks is more rigorous. As shown in Fig. 1, every filler item can be rated by one or two attackers in the group. IØ is the set of items that are rated by attackers in the attack group. IT is the set of target items whose ratings are determined by Function Depending on the purpose of attack, the target items are assigned to rmax (maximum rating) for push attacks or rmin (minimum rating) for nuke attacks. In , a generative model of group shilling attacks was proposed, which could be used to create group attack profiles with high diversity. This model includes two versions, i.e., a strict version and a loose version, which are denoted as GSAGens and GSAGenl, respectively. GSAGens (GSAGenl) has two types, i.e., GSAGens Ran and GSAGens Avg (GSAGenl Ran and GSAGens Avg), which are developed on the basis of random attack model and average attack model, respectively. Table I summarizes these group attack models. In Table I, GSAGens has more rigorous conditions in generating group attack profiles, so the size of the attack groups generated by this group attack model is limited. To ensure the effect of group shilling attacks, we use the loose version to generate group attack profiles in this article.



III. Existing System:

To protect recommender systems, various approaches have been presented to detect shilling attacks over the past decade. However, these approaches focus mainly on detecting individual attackers in recommender systems and rarely consider the collusive shilling behaviors among attackers. Although some approaches have been proposed to detect shilling behaviors at the group level, they divide candidate groups and identify attack groups according to profile similarity. There are some group attack models that can generate attack profiles with great diversity. As a result, these approaches cannot fully detect attack groups, which cause poor precision and recall. Recently, some approaches have been presented to detect spammer groups in review websites. However, the group shilling attacks in recommender systems are different from the spammer groups in review websites. Therefore, the spammer group detection approaches are not applicable to the detection of group shilling attacks.

IV. Proposed System

To overcome the abovementioned limitations, we propose a method to detect group shilling attacks in online recommender systems through dB scan . 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. The major contributions of this article are listed as follows.

1) 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.

2) 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, dB scan algorithm is employed to cluster the candidate groups according to their suspicious degrees, and the attack groups are obtained.

3) 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.

V. Implementation

Divide Candidate Groups:

- For each item i, find the users who rate item i and the corresponding rating time from the data set and then arrange them

in chronological order to construct the rating track of item i.

- In the rating track of item i, the first user's rating time is obtained and set as a starting point, thereafter extract users whose rating time is within TIL days after the starting point, and divide these users into a candidate group.
- In the rating track of item i, the rating time of the first user who is not in groups is selected as the new starting point, and then, the users whose rating time is within TIL days after the new starting point are extracted and divided into a candidate group.
- Repeat 3) until all users in the rating track of item i are divided into candidate groups.
- Repeat steps 1)–4) until all items are processed.

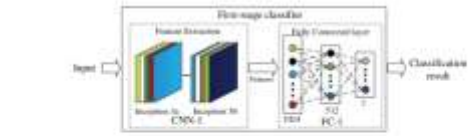
Calculate the suspicious degree of each candidate group:

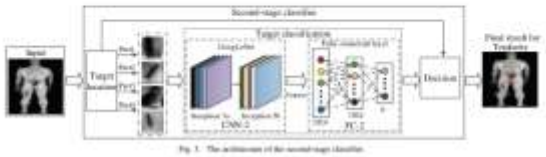
From the item perspective, the intent of an attack group is to increase the recommended probability of the target item. If attackers collude to promote or demote an item, the item's attention degree will be high. To achieve the desired attack effect, attackers in an attack group are required to complete their rating tasks within a specified time, so the attackers in the group will be active in this time interval. Therefore, if users in a group rate items with high attention degree and these users are active at the same time, the group is more likely to be an attack group. Based on these findings, the user features and item features are extracted to calculate the suspicious degrees of the candidate groups.

Detect attack groups:

Based on the divided candidate groups, we employ the dB scan algorithm to cluster the candidate groups according to their suspicious degrees and identify the attack groups from the generated clusters of candidate groups. More specifically, we take the set of GSDs as data samples and perform the dB scan on it. After clusters of candidate groups are generated, we calculate the mean of GSDs for each of the clusters. If the mean value for a cluster is greater than or equal to the sum of the average suspicious degree of the candidate groups and the standard deviation of the suspicious degrees of the candidate groups, the groups in this cluster are treated as the attack groups.

VI. System Architecture





VII. EXPERIMENTAL EVALUATION

A. Experimental Data Sets and Setting

To e valuate the proposed method (GD-BKM), the following two data sets are used to conduct experiments.

1)Netflix Data Set There are 103 297 638 ratings from 480 186 users on 17 770 movies in the data set. The ratings are integers between 1 and 5, where 1 and 5 indicate disliked and most liked, respectively. We randomly sample 215 884 ratings from 2000 users on 3985 movies as the experimental data set.

2) Amazon Review Data Set: This data set was constructed by Xu et al. including 1 205 125 comments/ratings from 645 072 users on 136 785 products. The ratings in the data set are integers between 1 and 5, where 1 and 5 indicate disliked and most liked, respectively. We extracted 5055 labeled users and their ratings on products and generated a labeled data set consisting of 53 777 ratings from 5055 users on 17 610 products. There are 1937 attackers and 3018 genuine users among the 5055 users. The average Pearson similarity between genuine users is about 0.0055, and the average Pearson similarity between attackers is about 0.0171. The experiment in this article consists of two parts. The first part of the experiment is conducted on a synthetic data set. We suppose that all users in the sampled Netflix data set are genuine users. The attack groups are generated with GSAGenl Ran and GSAGenl Avg models. The input of GSAGenl Ran and GSAGenl Avg is a set of attack profiles that are generated with a random attack model and an average attack model, respectively. The size of the generated attack group depends on the attack size and filler size used in the random and averageattack models.The smaller the filler size, the larger the group size, and the larger the attack size, the larger the group size [10]. However, the larger attack size requires a higher cost of attack. After many experiments of generating attack groups, we set the attack size and filler size of the random and averageattackmodelsto 10% and2.5%,respectively,toensure the size of the attack group. A total of ten attack groups are generated and injected into the sampled Netflix data set. The target item to be promoted by each attack group is randomly chosen from unpopular items in the data set. The rating time of attackers is randomly chosen within the continuous 30 days between the earliest and the latest time of the rated item. The average Pearson similarity between genuine users is about 0.1001, and the average Pearson similarity between attackers is about 0.008. In the second part of the experiment, we identify the attack groups in the Amazon data set and compare the experimental results of GD-BKM with those of baseline methods.

B. Evaluation Metrics

Precision and recall metrics are used to evaluate the performance of GD-BKM, which are defined as follows:

where TP denotes the number of attackers correctly identified, FN denotes the number of

attackers misjudged as genuine users, and FP denotes the number of genuine users misjudged as attackers.

We also use Precision and Recall metrics for evaluating the performance of GD-BKM, which are defined as follows [24]: Precision = |Uk ∩UA| |Uk| (7) Recall= |Uk ∩UA| |UA| (8) where Uk is the set of top-k users ranked according to the user suspicious degrees, UA is the set of all attackers in the data set, and |Uk ∩UA| represents the number of attackers in Uk and |Uk|=k.

C. Experimental Results and Analysis

- To show the effectiveness of GD-BKM, we compare GD-BKM with the following four methods.
- Catch the Black Sheep (CBS) [24]: An approach for detecting shilling attacks, which uses spam probability value to rank users. In this approach, a small number of attackers need to be labeled as the initial seed users. In our experiments, we randomly select five labeled attackers in the Amazon data set and 27 attackers in the Netflix data set as the seed users, respectively.
- 2) UD-HMM [25]: An unsupervised method for shilling attack detection, which models the behavior difference between attackers and genuine users using the hidden Markov models and employs Ward's hierarchical clustering to identify attackers. In the experiments, parameters N and α for UD-HMM are set to 5 and 0.8 on the Netflix data set and 15 and 0.7 on the Amazon data set, respectively.
- 3) DPTS [28]: A shilling attack detection method in which the item rating-time series is first dynamically partitioned based on important points and then the chi-square distribution ($\chi 2$) is used to detect abnormal intervals.
- 4) GSBC [23]: A method to discover spammer groups based on the reviewer graph, which divides candidate groups using the minimum cut algorithm. The candidate groups are ranked by their spam scores, and a threshold is given to judge whether a group is a spammer group. For the Amazon data set, we set τ = 30, δ = 0.1, MAXSIZE=50, and MINSPAM=0.49. For the Netflix data set, we set τ = 30, δ = 0.1, MAXSIZE = 50, and MINSPAM = 0. 1) Selection : In this section, we utilize the elbow rule to determine the value of dB scan a. In the elbow rule, the value is selected according to the sum of the squared errors (SSE). The main idea of the elbow rule is that there exists a point before which the SSE will decrease sharply with the increase. After that point, SSE will gradually flatten as the value continues to increase. In this case, the dividing point is the suitable value . SSE is calculated

VIII. 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 dB scan algorithm. The proposed detection model can overcome the problem that the performance is poor when attackers have a few corated 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. We combine the features of items and users to calculate the GSDs. Based on the GSDs, dB scan algorithm is utilized to identify attack groups from the candidate groups. The experimental results on two data sets illustrate the effectiveness of our method.

IX. 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 recommendation 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.

Vol 12, Issue 05, MAY /2021
ISSN NO:0377-9254

Conf. Adv. Data Min. Appl., Nanjing, China, 2012, pp. 675–688.

[11] K. Murugesan and J. Zhang, "Hybrid dB scan 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 Comput. 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.