IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

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

Mrs. M.Menakapriya, S. Ilayaragavan, G. Indira, M.Karthikeyan, K.M.Kalaivani,

Associate Professor, Department of MCA, Muthayammal Engineering College, Namakkal, Tamilnadu, India

Final Year Student, Department Of MCA, Muthayammal Engineering College, Namakkal, Tamilnadu, India

ABSTRACT: To protect recommender systems against shilling attacks, a variety of detection methods have been proposed over the past decade. However, these methods focus mainly on individual features and rarely consider the lockstep behaviours among attack users, which suffer from low precision in detecting group shilling attacks. In this work, we propose a three-stage detection method based on strong lockstep behaviours among group members and group behaviour features for detecting group shilling attacks. First, we construct a weighted user relationship graph by combining direct and indirect collusive degrees between users. Second, we find all dense subgraphs in the user relationship graph to generate a set of suspicious groups by introducing a topological potential method. Finally, we use a clustering method to detect shilling groups by extracting group behaviour features. Extensive experiments on the Netflix and sampled Amazon review datasets show that the proposed approach is effective for detecting group shilling attacks in recommender systems, and the $F1$-measure on two datasets can reach over 99 percent and 76 percent, respectively.

KEYWORDS: Recommender systems, Feature extraction, Clustering algorithms, Principal component analysis

I. INTRODUCTION

Information resources are growing explosively with the rapid development of Internet, which causes information overload. To deal with this problem, recommender systems have been widely used in e-commerce, social network, cloud computing, etc. [1]. However, due to the openness, malicious users can register accounts and provide ratings for the items with the intention of promoting or demoting the recommendation of target items. This behaviour has been termed as shilling attacks or profile injection attacks [2, 3]. Previous researches have shown that a large number of injected fake profiles can bias the output of collaborative filtering recommender systems, indicating that the collaborative filtering recommender systems are vulnerable to shilling attacks [4, 5]. Over the past decade, traditional attack models have been well studied, such as random attack, average attack, bandwagon attack, AoP attack, Love/hate attack, etc. In these attacks, attack users try to promote or demote the same target item but select filler items separately. In fact, attack users can easily collude to attack a set of target items in the recommender system. The behaviour wherein a group of attack users collude to promote or demote the recommendation of a set of target items has been termed group shilling attacks [6, 7]. As attack users in the group work together and strategically generate attack profiles, the detection methods proposed for detecting traditional attacks become ineffective for group shilling attacks. Therefore, how to improve the detection performance for group shilling attacks has become a key issue in the recommender systems.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

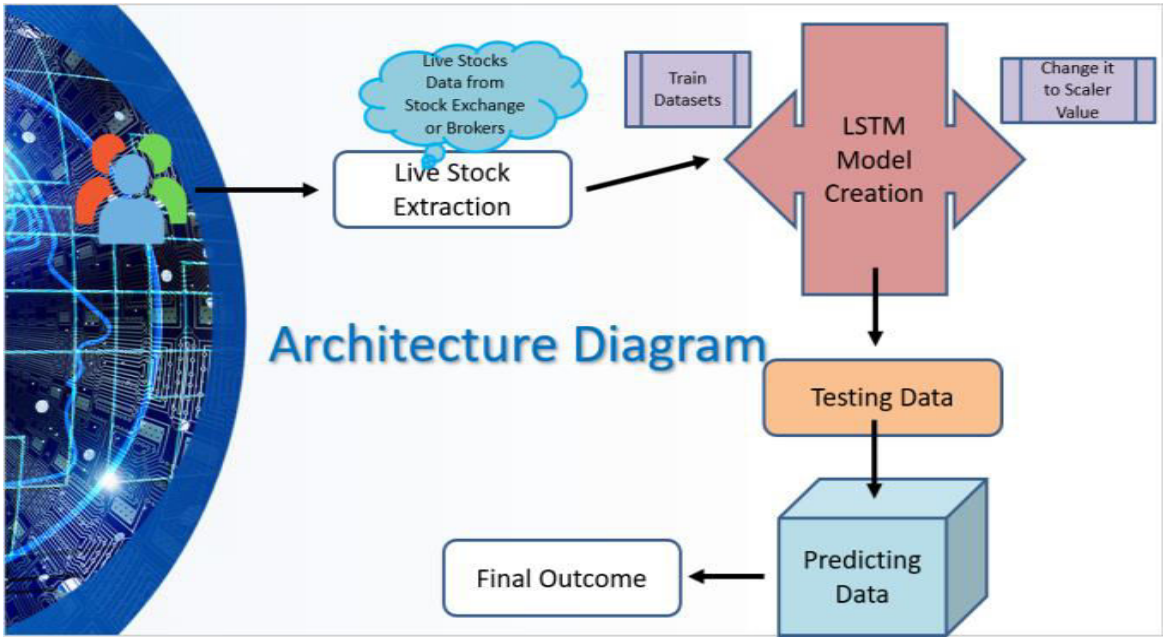

Fig 1: Architecture Diagram

To reduce the impact of shilling attacks on recommender systems, researchers have developed a variety of shilling attack detection methods [8–31]. Most methods focus mainly on the features of individual profiles and are used to detect attack profiles for traditional attacks, e.g., random attack and bandwagon attack. These methods rarely consider the collusive behaviours among attack users, which cause low precision in detecting group shilling attacks. Although a few detection methods have been proposed for detecting group shilling attacks in recent years, they usually need prior knowledge of attacks (e.g., the number of shilling groups or attackers). Otherwise, they suffer from low detection precision. While some approaches have been presented to spot the collusive groups in e-commerce and social network, the behaviour features of shilling groups vary greatly in different fields. Therefore, the approaches developed for identifying collusive groups in other fields are not suitable for detecting shilling groups in recommender systems.

To address the above limitations, we propose an unsupervised method for detecting group shilling attacks in recommender systems based on topological potential theory and group behaviour features, which is named as TP-GBF. The proposed approach focuses on the collusive behaviours among attack users and can accurately detect attack users in group shilling attacks without knowing the number of shilling groups in advance. Particularly, we first construct a weighted user relationship graph by analysing the lockstep behaviours between users. In the graph, vertices and edges represent the users and their relations, respectively. The weight of each edge is calculated by combining direct and indirect collusive degrees. Then, we analyse the network structure of shilling groups and introduce a topological potential method to spot suspicious groups. Finally, we use the clustering method to detect shilling groups by extracting group behaviour features.

II. BACKGROUND AND RELATED WORK

In Ref. [6], Su et al. first proposed the concept of group shilling attacks in recommender systems and mentioned two attack scenarios. In these scenarios, multiple attackers are well organized to conceal their intentions. The attackers in a shilling group work together to promote or demote a set of target items in scenario 1 or only attack some items in the target set in scenario 2. In order to achieve the attack intention, a shilling group must contain a certain number of attackers and each attacker in the group should give some normal ratings on nontarget items. Moreover, many shilling groups may coexist in a recommender system.

To generate the effective group shilling profiles and avoid the detection of the existing methods, Wang et al. [7] proposed a tricky group shilling attack model which includes a strict version denoted as $GSAGen_s$ and a loose version denoted as $GSAGen_l$. In their model, some shilling profiles generated by standard attack models, i.e., random attacks or average attacks, are used as the input. Then, they construct high diversity shilling profiles based on the input according to the strict condition or loose condition. Of the two versions, $GSAGen_l$ can generate a shilling group with more shilling profiles than $GSAGen_s$ because $GSAGen_l$ has less strict conditions in generating group shilling attack profiles. Table 1 summarizes the group shilling attack models.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

As for supervised detection methods of shilling attacks in recommender systems, Chirita et al. [8] first proposed the features named rating deviation from mean agreement (RDMA) and degree of similarity with top neighbours (DegSim). The effectiveness of two features depends on the attack type, attack size, and filler size. Burke et al. [9] trained a classifier based on some generic and model-specific features, which were extracted by utilizing statistics properties of ratings given by attack users. Although these detection features are helpful for detecting standard attacks, they become ineffective under obfuscated attacks. Wu et al. [10] presented an effective feature selection method for detecting five attacks and proposed the detection algorithm based on supervised learning. However, the proposed method needs to know the type of attack in advance. Yang et al. [11] proposed an ensemble shilling attack detection method based on 18 statistical features, which could obtain better detection performance than the baseline methods using a single classifier. However, it needs to consider the intensive numeric calculation on these features. Tong et al. [12] applied convolutional neural network to extract detection features from user-item rating matrix. This method can detect shilling profiles generated by the single attack method, but it is not suitable for detecting hybrid attacks. Hao et al. [13] first extracted multidimensional detection features from different angles, including ratings, rating time, and item popularity degree, and then they trained a SVM-based classifier for detecting shilling attacks. This method illustrates higher effectiveness than the baseline methods, but the detection performance is not good when the attack size is less than 10%.

III. METHODS

The Framework of TP-GBF

The framework of TP-GBF is depicted in Figure 1, which consists of three stages, i.e., constructing a weighted user relationship graph, generating a set of suspicious groups, and finding shilling groups. In the first stage, we construct a weighted user relationship graph by combining direct and indirect collusive degrees between users. The aim was to highlight the collusive behaviours among attackers in the same shilling group. In the second stage, we introduce a topology potential method to generate the set of suspicious shilling groups. The number of suspicious groups is decided automatically based on the number of nodes with local maximum potential value. In the third stage, group behaviour features are extracted at the group level, and then shilling groups in the set of suspicious groups are detected by using a clustering algorithm.

Many features have been proposed based on the individual attack profiles over the past decade. However, the effective detection features for shilling groups are limited. Since there may be a strong lockstep relationship between some normal users, groups of these normal users may be misjudged as shilling groups. In this section, six group behaviour features are proposed at the group level to distinguish shilling groups from normal ones in the set of suspicious groups.

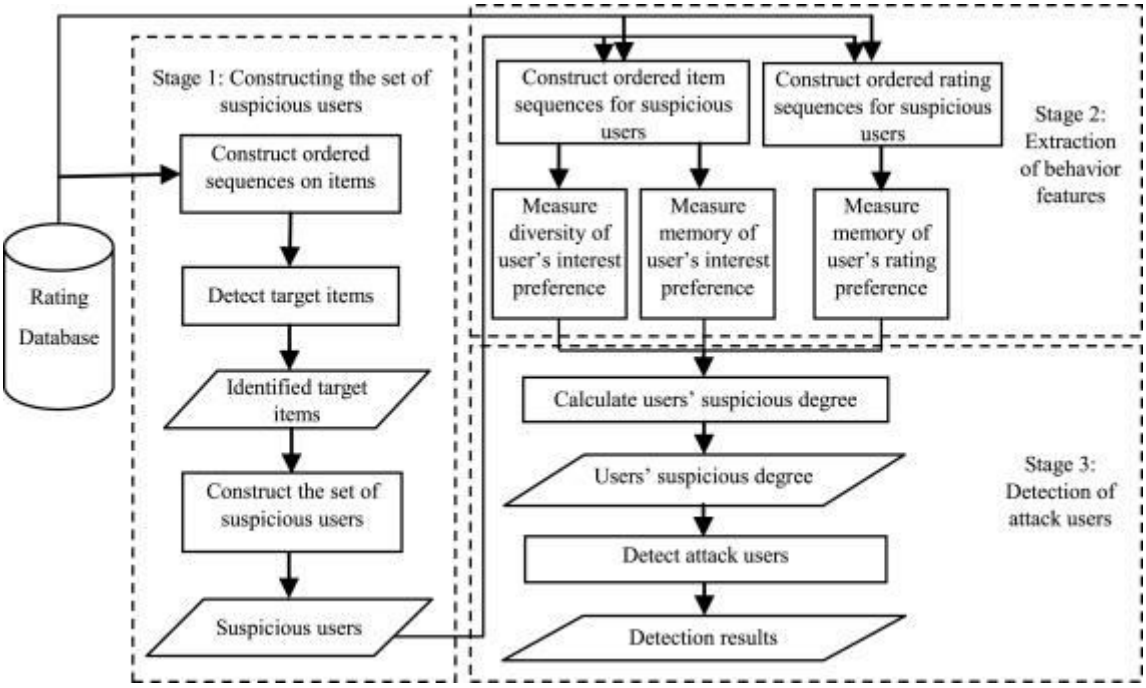

Fig 2: Work Flow

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

Semi-supervised detection methods first train weak classifiers with a few labeled user profiles. Then, some unlabeled user profiles are used to improve the weak classifiers. This type of method reduces the dependence on prior knowledge and assumptions. However, existing semi-supervised detection methods need extract hand-designed features of recommendation attack which is a challenging task even for domain experts. Supervised detection methods usually have good detection performance by training the classifiers with labeled user profiles. They learn knowledge from the training samples instead of depending on prior knowledge or assumptions for the detection. In particular, the recently proposed deep learning-based detection methods [23–27] do not need the hand-designed features and can automatically learn the features of recommendation attack besides having excellent detection performance. In the deep learning-based detection methods, there are usually many hyperparameters need to be set. These hyperparameters can greatly affect the detection performance of the detection methods. Although, some of the hyperparameters such as learning rate, loss function, and so on can be well determined by referring to relevant research results, the remaining hyperparameters such as activation function, epochs, and so on, which are called key hyperparameters for ease of discussion, need to be optimized differently for different CRS. In most of the existing deep learning-based detection methods, however, the key hyperparameters are optimized by manual analysis. That is, domain experts are employed to determine the key hyperparameters by analyzing the detection performance of candidate solutions. This way of determining key hyperparameters relies too much on domain experts and their experience. The swarm intelligence optimization algorithms have been effectively applied in intrusion detection (ID) for many years . For example, the Genetic Algorithm (GA) is used to optimize features and parameters of the classifier which is used to identify attacks in the ID systems. The Particle Swarm Optimization (PSO) algorithm is combined with some machine learning algorithms, such as k-means, to improve the performance of anomaly detection. The Ant Colony Optimization (ACO) algorithm is combined with Decision Tree to build a multiple level hybrid classifier for classifying attacks in the ID systems.

IV. RESULT ANALYSIS

When a group of attackers join forces to take down the recommendation systems, they rate not only the targeted item(s), as well as some non-target items. Furthermore, assailants in order to accomplish the intended assault impact, the group must complete their rating tasks within a specified amount of time. On the basis of these assumptions, the above diagram presents a group shilling attack detection method based on bisecting K-means clustering and hierarchical clustering. The propose diagram consists of three phases. The first stage is to create candidate groups, which are made up of individuals that score the same thing at the same time. In the second stage, the user and item characteristics are retrieved. The degree of suspicion for each application group is determined by combining these criteria.

The system clusters the candidates are divided into categories based on their level of suspicion. using bisecting K-means and hierarchical clustering and based on the separated candidate groups, determine the assault groups from the produced groupings of candidate groups. In more detail, it uses the collection of GSDs (Group Suspicious Degrees) as data samples, then use K-means clustering to bisect them and grouping in a hierarchical manner between them. This establishes the mean of GSDs for each of the K clusters after generating K clusters of groupings of candidates. If the cluster's mean value is more than or equivalent to the mean suspicious degrees of the candidate organizations plus overall statistical significance of the candidates groups' suspicious degrees, the groupings in this cluster are considered attack groups. The number of clusters K is necessary in k-means clustering, however with hierarchical clustering, no previous knowledge of the clustering is necessary.

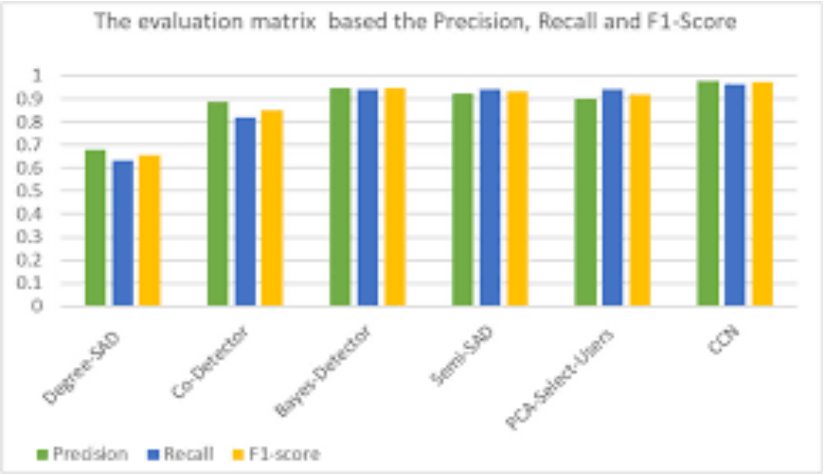

Fig 3: Group Shilling Attacks Result Analysis

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

Shilling assaults in groups pose a significant danger to recommender systems. The bisecting K-means-based group assault detection model and hierarchical clustering technique to detect such assaults. When attackers have a few co-rated items, the suggested detection approach can solve the problem of low performance. The beginning time point for dividing every item's rating track to split candidate groups is dynamically determined using a set time duration. The features of objects and users are mixed to compute the GSD. The bisecting K-means technique is used to separate candidate groups are being attacked using GSDs. Our technique's success is demonstrated by the results of our testing on Amazon data sets. When the methods bisecting K-means and hierarchical clustering are compared, the hierarchical clustering algorithm outperforms the bisecting K-means clustering algorithm.

V. CONCLUSIONS

Compared with traditional shilling attacks in recommender systems, group shilling attacks are more harmful and much harder to detect. To improve the detection performance and minimize the impact of group shilling attacks on recommender systems, we present a three-stage group shilling attack detection method based on lockstep behaviour and group behaviour features. We propose a method of calculating collusive degree between users based on lockstep behaviour and transitivity of direct relations, based on which a weighted user relationship graph is constructed. We calculate the topology potential for each user node in the weighted user relationship graph and find out all suspicious groups by local maximum potential nodes. We propose six group behaviour features to characterize the differences between shilling groups and normal ones in the set of suspicious groups, and devise a k-means clustering-based algorithm to detect the shilling groups. The experimental results on the Netflix and Amazon datasets show that TP-GBF is effective for detecting group shilling attacks in recommender systems and outperforms three baseline methods in precision, recall, and F1-measure.

REFERENCES

1. F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, Springer, Berlin, Germany, 2015.
2. M. Si and Q. Li, "Shilling attacks against collaborative recommender systems: a review," *Artificial Intelligence Review*, vol. 53, no. 1, pp. 291–319, 2020.
   View at: Publisher Site | Google Scholar
3. Y. Wang, L. Qian, F. Li, and L. Zhang, "A Comparative study on shilling detection methods for trustworthy recommendations," *Journal of Systems Science and Systems Engineering*, vol. 27, no. 4, pp. 458–478, 2018.
   View at: Publisher Site | Google Scholar
4. I. Gunes, C. Kaleli, A. Bilge, and H. Polat, "Shilling attacks against recommender systems: a comprehensive survey," *Artificial Intelligence Review*, vol. 42, no. 4, pp. 767–799, 2014.
   View at: Publisher Site | Google Scholar
5. A. M. Turk and A. Bilge, "Robustness analysis of multi-criteria collaborative filtering algorithms against shilling attacks," *Expert Systems with Applications*, vol. 115, pp. 386–402, 2019.
   View at: Publisher Site | Google Scholar
6. X. Su, H. Zeng, and Z. Chen, "Finding group shilling in recommendation system," in *Proceedings of the Special Interest Tracks & Posters of the International Conference on World Wide Web*, pp. 960-961, New York, NY, USA, May 2005.
   View at: Publisher Site | Google Scholar
7. Y. Wang, Z. Wu, J. Cao, and C. Fang, "Towards a tricksy group shilling attack model against recommender systems," *Advanced Data Mining and Applications*, vol. 7713, pp. 675–688, 2012.
   View at: Publisher Site | Google Scholar
8. P. Chirita, l. W. Nejd, and C. Zamfir, "Preventing shilling attacks in online recommender systems," in *Proceedings of the 7th annual ACM international workshop on web information and data management*, pp. 67–74, Bremen Germany, November 2005.
   View at: Publisher Site | Google Scholar
9. R. Burke, B. Mobasher, and C. Williams, "Classification features for attack detection in collaborative recommendation systems," in *Proceedifgs of the 12th International Conference on Knowledge Discovery and Data Mining*, pp. 542–547, Philadelphia, PA, USA, August 2006.
   View at: Publisher Site | Google Scholar
10. Z. Wu, Y. Zhuang, and Y. Wang, "Shilling attack detection based on feature selection for recommendation system," *Journal.Acta Electronica Sinica*, vol. 40, no. 8, pp. 1687–1693, 2012, in Chinese with English abstract.
    View at: Google Scholar

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 4, April 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1204056 |

11. Z. Yang, L. Xu, Z. Cai, and Z. Xu, "Re-scale AdaBoost for attack detection in collaborative filtering recommender systems," *Knowledge-Based Systems*, vol. 100, pp. 74–88, 2015.
View at: Publisher Site | Google Scholar
12. C. Tong, X. Yin, J. Li et al., "A shilling attack detector based on convolutional neural network for collaborative recommender system in social aware network," *The Computer Journal*, vol. 61, no. 7, pp. 949–958, 2018.
View at: Publisher Site | Google Scholar
13. Y. Hao, P. Zhang, and F. Zhang, "Multiview ensemble method for detecting shilling attacks in collaborative recommender systems," *Security and Communication Networks*, vol. 2018, no. 4, Article ID 8174603, 33 pages, 2018.
View at: Publisher Site | Google Scholar
14. Q. Zhou, J. Wu, and L. Duan, "Recommendation attack detection based on deep learning," *Journal of Information Security and Applications*, vol. 52, Article ID 102493, 2020.
View at: Publisher Site | Google Scholar
15. S. Zhang, H. Yin, and T. Chen, "GCN-Based representation learning for unifying robust recommendation and fraudster detection," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 689–698, Xi'an, China, July 2020.
View at: Publisher Site | Google Scholar

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details