Defending Suspected Users by Utilizing Specific Distance Metric in Collaborative Filtering Recommender Systems

Prof. Vinod Bharat[#1,] Sanket Sonu[#2] , Phadtare Omkar Gajanan[#3] , Siddhesh Salunkhe[#4]

[1]*Head of Computer Department, D Y Patil School of Engineering Academy,Ambi, Pune, Maharashtra, India*
[#]*Computer Science Department,D Y Patil School Of Engineering Academy,Ambi , Pune - 410507*

Abstract— **Recommender system is an imperative part of the data and internet business biological system. Collaborative filtering (CF) is a vital and well known innovation for recommender system. Collaborative filtering recommender systems (CFRSs) are basic parts of existing well known online business sites to make personalized recommendations. Be that as it may, current CF strategies experience the ill effects of such issues as "shilling" attacks or "profile injection" attacks because of its openness. While an extensive variety of recognition systems have been utilized, some of them depended on calculating similarity between users (consists of attackers and genuine users) keeping in mind the end goal to segregate those attackers. In practice, it is hard to catch every concerned attacker by abusing the similarity of users, in spite of the fact that it can be useful to filtering out more genuine users. Calculating directly the similarity between users on a whole dataset consumed high computation time although they can be powerful to catch the concerned attackers in some degree. In this paper we propose an unsupervised method to detect such attacks. At the first stage, we filter out a part of genuine users in order to reduce the enumeration time. At the second stage, we mainly focus on the effective similarity metric to better differ between attackers and genuine users based on the remaining users of the first stage and modulates the traditional similarity metric and the linkage information between users to improve the accuracy of similarity of users.**

Keywords— **Recommender system, Shilling attack, Attack detection, Collaborative Filtering (CF), Similarity metric.**

I. INTRODUCTION

Personalization recommender systems (RSs) become more and more in demand in e-commerce websites to automatically make personalized suggestions of services or products to clients. Collaborative filtering (CF) is an essential and well known technology for recommender systems. There has been a lot of work done both in industry and the scholarly community. These methods are classified into user-based CF and item-based CF. The basic idea of user-based CF method is to search out a set of users who have similar favor patterns to a given user (i.e., "neighbors" of the user) and recommend to the user those items that other users in the same set like, while the item-based CF method aims to confer a user with the recommendation on an item based on the other items with high correlations (i.e., "neighbors" of the item). In all collaborative filtering methods, it is a valuable step to find users' (or items') neighbors, that is, a set of similar users (or items). However, Collaborative filtering recommender systems (CFRSs) are prone to manipulation from attackers due to its openness, which carefully infuse chosen attack profiles into CFRSs to prejudice the recommendation results to their benefits or decrease the trustworthiness of recommendation. These events are often called "shilling" attacks or "profile injection" attacks. Therefore, constructing an effective detection method to detect the attackers and remove them from the CFRSs is vital. A number of detection methods have been proposed to make CFRSs obstruct to such attacks. Some of them were based on calculating similarity between users (consists of attackers and genuine users) in order to differentiate those attackers. In practice it is hard to catch all concerned attackers by exploiting the similarity of users, albeit it can be helpful to filtering out more genuine users and show high computation time, albeit they can be effective to capture the concerned attackers in some extent. To resolve all the issues in existing systems we propose a new detection method called 0 suspected users by exploiting specific distance metric in collaborative filtering recommender systems to make CFRSs resistant to such attacks, which exploits a novel metric for calculating similarity between users. To detract the computation time, we firstly filter out a part of genuine users in order to detract the computation time. Since the attackers will target one or more specific items with lowest or highest rating many times if they want to demote (called nuke attack) or promote (called push attack) the items to the recommendation list, we can find out all suspected target items by using an complete count threshold. At the second stage, we primarily concentrate on the effective similarity metric to better distinguish between attackers and genuine users based on the remaining users of the first stage.

II. LITERATURE SURVEY

C. Chung, P. Hsu, and S. Huang, [1] presented a novel approach to filter out malicious rating profiles from recommender systems. Authors propose a novel detection method to make recommender systems resistant to the "shilling" attacks or "profile injection" attacks. They firstly work out the problem as finding a mapping model between rating behavior and item distribution by exploiting the least-squares approximate solution. Based on the trained model, they plan a detector by employing a regress or to detect such attacks.

F. Zhang and Q. Zhou, [2] proposed an online method for detecting profile injection attacks in collaborative recommend systems. Authors propose an online method (called HHT–SVM) to detect profile injection attacks by combining Hilbert–Huang transform (HHT) and support vector machine (SVM), which can work incrementally. The underpinning idea of HHT–SVM is the feature extraction method based on an individual user profile. They first build rating series for each user profile based on the novelty and popularity of items. Then, by familiarizing HHT they use the empirical mode decomposition (EMD) approach to decompose each rating series and extract Hilbert spectrum based features to characterize the profile injection attacks. Finally, they exploit SVM to detect profile injection attacks based on the proposed features.

W. Zhou, Y. S. Koh, J. H. Wen, S. Burki, and G. Dobbie, [3] proposed detection of abnormal profiles on group attacks in recommender systems which focuses on searching shilling groups from Amazon China, a typical real e-commerce website. To this end, a set of general arrangements are proposed, though the ground truth labels are unknown. Specifically, they first employ the FIM technique to generate candidate groups, and then present several features for modeling the abnormal behavior at the group level. By transforming each group into a feature space, a ranking method called PCA-SGD is designed for unsupervised detection.

M. Morid and M. Shajari, [4] demonstrates an attack detection method based on user influence in recommender systems. Influential users from a recommender system user set were recognized and used to extract detection features from each user profile. A KNN data mining model that uses quoted features as input was applied to the selected user set to identify user profiles that were from attackers as its output.

Z. Zhang and S. R. Kulkarni,[5] proposed detection of shilling attacks in recommender systems via spectral clustering which define the issue as finding a most extreme submatrix in the user-user similarity matrix. To find a most extreme submatrix, authors translate the matrix into a graph and apply a spectral clustering algorithm to find the min-cut solution to estimate the highly correlated group. The graph is created based on the edge density in order to allow dealing with an unbalanced clustering.

I. Gunes, C. Kaleli, A. Bilge, and H. Polat, [6] presented shilling attacks against recommender systems: A comprehensive survey. Decoding shilling attack detection schemes in detail and robust algorithms proposed so far might open a lead to develop new detection schemes and increase such robust algorithms further, even propose new ones. Thus, authors describe various attack types and introduce new dimensions for attack classification. Detailed description of the proposed detection and robust recommendation algorithms are given.

Z. Zhang and S. Kulkarni, [7] proposed graph-based detection of shilling attacks in recommender systems. Authors present a method to make recommender systems obstructive to these attacks in the case that the attack profiles are highly correlated with each other. They plan the issue as finding a maximum submatrix in the similarity matrix. They search for the maximum submatrix by changing the problem into a graph and merging nodes by heuristic functions or finding the largest component.

Z. A. Wu, Y. Q. Wang, and J. Cao, [8] proposed a survey on shilling attack models and detection techniques for recommender systems. This paper reviews the states of art and the main problems of existing works related to shilling attack models and detection techniques, and attempts to sketch a extensive and evident outline for this new and active research realm. In particular, the motivations, concepts, intent, ingredients and classifications of the shilling profiles are introduced, and two kinds of metrics for evaluating the harmness of shilling attacks are presented.

C. E. Seminario and D. C. Wilson., [9] presented attacking item-based recommender systems with power items that uses influential items to successfully attack RSs. They show that the Power Item Attack (PIA) is ready to affect not only user-based and SVD-based recommenders but also the heretofore highly robust item-based approach, using a novel multi-target attack vector.

J. Hu, D. C. Zhan, X. Wu, Y. Jiang, and Z. H. Zhou,[10] presented pair wised specific distance learning from physical linkages. This approach exploits the structures of physical linkages and in particular captures the key observations that nonmetric and clique linkages imply the appearance of different or unique semantics, respectively. It is notable that, rather than generating a global distance, PSD generates different distances for different pairs of data points; this property is desired in applications involving complicated data semantics. They mainly present PSD for multi-class learning and further extend it to multi-label learning.

III. Existing System Approach

Constructing an effective detection method to detect the attackers and remove them from the CFRSs is crucial. In the existing systems a number of detection methods have been proposed to make CFRSs resistant to such “shilling” attacks or “profile injection” attacks. Some of them were based on calculating similarity between users (consists of attackers and genuine users) in order to distinguish those attackers. Practically speaking it is hard to catch all concerned attackers by exploiting the similarity of users, although it can be helpful to filtering out more genuine users and attackers and show high computation time, although these existing systems can be effective to capture the concerned attackers in some extent.

IV. Proposed System Approach

In this Paper, we propose a new detection method to make CFRSs resistant to such attacks, which exploits a novel metric for calculating similarity between users. To lower the computation time, firstly we filter out more genuine users as far as possible determined by using mistrusted target items. Since the attackers will target one or more specific items with lowest or highest rating many times if they want to demote (called nuke attack) or promote (called push attack) the items to the recommendation list, we can find out all mistrusted target items by using an complete count threshold. Based on the remaining users, we employ a new similarity metric inspired from the pair wised specific distance, directing to measure impressively the similarity between users.

V. Proposed Architecture

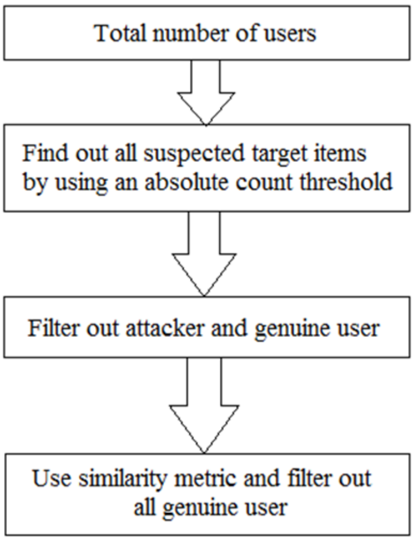
Total number of users
Find out all suspected target items by using an absolute count threshold
Filter out attacker and genuine user
Use similarity metric and filter out all genuine user

Figure 1: System Architechture

Our proposed approach consists of two stages: the stage of filtering out genuine users by misusing suspected target items and the stage of separating attackers by employing a new similarity metric. At the first stage, we filter out a part of genuine users in order to decrease the computation time. To decide the target items, we design an absolute count threshold ε which pushes or nukes the same items with the highest or lowest at least ε times. If the count for an item is greater than ε, then the item *susi* is regarded as suspected target item. Users (consist of attackers and genuine users) who rated the *susi* with the highest *rmax* or lowest *rmin* are considered as attackers. It is noticeable that there will be more false positives or false negative if ε is too small or large. At the second stage, we mainly focus on the effective similarity metric to better differentiate between attackers and genuine users based on the remaining users of the first stage.

VI. Conclusion

In this paper, we have exhibited an unsupervised detection method called defending suspected users by exploiting specific distance metric in collaborative filtering recommender systems for detecting attacks like “shilling” attacks or “profile injection” attacks, which exploits the pair wised specific distance to generate a similarity metric. Firstly, we filter our more genuine users as far as possible determined by using mistrusted target items. And then, based on the remained users, we continue to filter out more genuine users used by an empirical threshold of the new similarity metric.

Acknowledgment

The research presented in this paper is supported in part by the National Natural Science Foundation (61221063, U1301254), 863 High Tech Development Plan (2012AA011003) and 111 International Collaboration Program, of China.

References

[1] C. Chung, P. Hsu, and S. Huang, “A novel approach to filter out malicious rating profiles from recommender systems,” *Journal of Decision Support Systems*, pp. 314–325, 2013.

[2] F. Zhang and Q. Zhou, “HHT-SVM: An online method for detecting profile injection attacks in collaborative recommender systems,” *Knowledge-Based Systems*, 2014.

[3] W. Zhou, Y. S. Koh, J. H. Wen, S. Burki, and G. Dobbie, “Detection of abnormal profiles on group attacks in recommender systems,” *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 955–958, 2014.

[4] M. Morid and M. Shajari, “Defending recommender systems by influence analysis,” *Information Retrieval*, pp. 137–152, 2014.

[5] Z. Zhang and S. R. Kulkarni, “Detection of shilling attacks in recommender systems via spectral clustering,” *International Conference on Information Fusion*, pp. 1–8, 2014.

[6] I. Gunes, C. Kaleli, A. Bilge, and H. Polat, “Shilling attacks against recommender systems: A comprehensive survey,” *Artificial Intelligence Review*, pp. 1–33, 2012.

[7] Z. Zhang and S. Kulkarni, “Graph-based detection of shilling attacks in recommender systems,” *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6, 2013.

[8] Z. A. Wu, Y. Q. Wang, and J. Cao, “A survey on shilling attack models and detection techniques for recommender systems,” *Science China*, vol. 59, no. 7, pp. 551–560, 2014.

[9] C. E. Seminario and D. C. Wilson., “Attacking item-based recommender systems with power items,” *ACM Conference on Recommender Systems*, pp. 57–64, 2014.

[10] J. Hu, D. C. Zhan, X. Wu, Y. Jiang, and Z. H. Zhou, “Pairwised specific distance learning from physical linkages,” *ACM Trans. Knowl. Discov. Data*, 2014.