2018 IEEE International Conference on Systems, Man, and Cybernetics

Shilling Attack Detection using Rated Item Correlation for Collaborative Filtering

Keke Chen
School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
Email: kkechen@qq.com

Patrick P. K. Chan
School of Computer Science and Engineering, South China University of Technology, Guangzhou, China
Email: patrickchan@ieee.org

Daniel S. Yeung
Systems, Man, and Cybernetics Society, United States
Email: danyeung@ieee.org

Abstract— **Collaborative filtering (CF) is vulnerable under shilling attack, which misleads recommendation of CF by injecting well-crafted profiles to a targeted system. Although a number of supervised learning based shilling attack detection methods are proposed, their features mainly measure rating values and items of a profile individually, but ignore the relation between items. This study aims to enhance the robustness of CF against shilling attack by considering the rated item correlation. Real users rate items based on their preferences, but rated items are randomly selected for malicious users profiles in most shilling attack. Therefore, the rated item correlation of real and malicious profiles is different. Three features are proposed to capture the information from different intervals of the distribution of rated item correlation in terms of Cosine Association (CA). A benchmark dataset, MovieLens 100K, is used to evaluate the proposed features. The discrimination ability of the proposed features is also illustrated. The experimental results suggest that the proposed features have significant contribution on shilling attack detection.**

I. Introduction

In the era of information explosion, the difficulty of choosing desired information increases as people are overwhelmed with data. Recommender systems play an important role to select suitable choices from plenty of options to users. Collaborative filtering (CF) [1] which is one of the most common recommendation systems, recommends preferences to a target user according to other users similar to the target user. CF achieves satisfying performance in numerous applications. However, some studies [2, 3] indicate that the vulnerability of CFs in an adversarial environment in which an adversary manipulates the dataset in order to mislead decisions of CF on purpose.

Shilling attack (or profile injection attack) [3] assumes a number of malicious user profiles are injected to the database of a recommender system to influence the chance of recommending a target item to users. Malicious user profiles are carefully crafted aiming to increase the similarity to the profiles of real users. A number of items are selected for each malicious profile, and a value is assigned to them according to the behaviours of real users. As traditional CF methods do not aware of the exist of shilling attack, their recommendation is changed dramatically.

A number of shilling attack detection methods have been proposed to distinguish malicious profiles from real user profiles. Unsupervised learning techniques [4] are used to cluster user profiles according to characteristics of the profiles in some studies. As no knowledge of shilling attack is required in advance, the accuracy of these methods drops easily when malicious profiles are similar to the real ones. On the other hand, assuming some information of attack is obtained, some attack detection adopt a supervised learning method [5], *i.e.,* a classifier learns from a dataset containing real and malicious user profiles. Different properties of real and malicious user profiles are measured by a number of features, named as generic attributes, to identify malicious profiles. In this study, the detection with supervised learning is focused since it is generally more accurate than the one with unsupervised learning.

With limited knowledge on a targeted system, most shilling attack strategies generate malicious user profiles by randomly choosing rated items. However, this rating behaviour is different from the one of genuine users who always rate an item according to their preferences, such as some items have higher chance to be rated together. For instance, a person who likes iWatch may also like iPhone rather than Windows Phone. The correlation of rated items is closely related to the detection accuracy of shilling attack detection. However, to the best of our knowledge, the current features used in supervised learning based shilling attack detection methods only consider similarity of a profile with its nearest neighbors [6], rating values in a profile [7], and rated item distribution [8]. All existing features measure a rated item in a profile individually, but not the correlation of the items.

By this observation, this study investigates whether and how rated item correlation improves the detection accuracy on shilling attacks. Cosine Association (CA) [9] is applied to measure the rated item correlation. Features based on CA are proposed in order to capture the difference of rated item correlation between malicious and real user profiles. The discrimination ability of the proposed features are illustrated experimentally. The experimental comparison between the supervised learning based detection method with and without

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00601

using the proposed features in different attack scenarios is then provided. The results confirm the contribution of the proposed feature to shilling attack detection.

The rest of the paper is arranged as follows. The related concepts of this study are introduced in Section II. The measurement of rated item correlation and the proposed method are described in Section III. Section IV illustrates and discusses the experimental results of our proposed and existing methods. Finally, the conclusion and future work are described in Section V.

II. Background

This section gives a brief introduction on the major concepts, including the shilling attack models and its countermeasure related to this study.

A. Shilling Attacks

A number of malicious user profiles are injected into a recommender system in shilling attack in order to mislead the frequency of recommending a target item. The attack is separated into two types based on its objective: *push attack* increases the frequency of recommending a target item, while the frequency reduces in *nuke attack* [10]. *Attack size* and *filler size*, which are the parameters of shilling attack, controls the number of malicious user profiles injected to a system and the number of items rated in each profile. Higher attack cost is required for an attack model with larger As or Fs.

A malicious user profile generated by a shilling attack model usually includes four parts: *target item* (I_t) is preselected by an adversary. It is rated with the maximum (minimum) value in the push (nuke) attack; the selected items (I_s) contains a set of rated items aiming to increase the similarity between the malicious and genuine profiles; the filler items (I_f) are randomly selected to increase the diversity of malicious user profiles; and the unrated items (I_u) are the rest of items.

Only I_f but not I_s is considered in both *Random* and *Average attack*. All the rated items are selected randomly. In contrast, *Average over Popular (AoP) x% attack* [11] neglects I_f, and I_s contains rated items randomly selected from the $x\%$ most popular items. On the other hand, *Bandwagon attack* considers both I_s and I_f in the generated profiles, and the I_s included some popular items. Although some other attacks exists, most of them are revised from these four standard attack models.

B. Shilling Attack Detection Methods

Shilling attack detection is formulated as an unsupervised and a supervised learning problem. Unsupervised learning methods, *e.g.,* K-mean, Probabilistic Latent Semantics Analysis (PLSA) and Principle Component Analysis (PCA) [12], separate user profiles into clusters by analyzing their characteristics. The minority of profiles is identified as shilling attack. Although no prior knowledge on shilling attack is required, the attack detection with unsupervised learning methods is easily evaded [13].

Some studies detect shilling attack by using supervised learning techniques, *e.g.,* Support Vector Machine (SVM), k-Nearest Neighbourhood (kNN) and C4.5. The main research problem in supervised learning based detection is to design a feature which is able to separate malicious and real user profiles accurately. The current features can be separated into three categories based on the type of captured information [14]. *Profile similarity* based features capture the similarity of the profiles and its nearest neighbors. For example, *Degree of Similarity with Top Neighbors* (DegSim) attribute calculates the average similarity of each profiles' k nearest neighbors [6]. The rating value of profiles is analyzed by *rating behavior* based features. *Rating Deviation from Mean Agreement* (RDMA), *Weighted Deviation from Mean Agreement* (WDMA), and *Weighted Degree of Agreement* (WDA) are the examples [7] . They calculate the deviation of rating score from the average rating score per item weighted by the inverse of the rated times of each item. *Item distribution* based features focus on which items are rated in a profile. For example, *Filler Size with Popular Items in Itself* (FSPII) calculates the ratio of between the number of popular items rated by the user and the total number of items rated by the user, and *Filler Size with Unpopular Items in Itself* (FSUII) calculates the proportion of unpopular items [8]. However, items of a profile is investigated individually by these features, and the correlation of items is neglected.

III. Proposed Method

Genuine users rate items according to their preference while rated items are usually selected randomly in existing shilling attacks. The rationale of our proposed method is to capture the rating behaviour of users in terms of rated item correlation to increase the accuracy of detection. CA is adopted as the measurement of rated item correlation to this study. In this section, CA is firstly introduced in Section III-A, and then the proposed feature is discussed in detail in Section III-B.

A. Cosine Association

CA is used as the measurement of item correlation in an association rule of data mining [9]. CA calculates the correlation between an item pair without considering their rating values, and indicates the possibility for a user to rate the item i and j at the same time. The definition of CA is shown as follows:

$$CA(i,j) = \sqrt{\frac{|U_{i,j}|^2}{|U_i||U_j|}} \tag{1}$$

where $|U_{i,j}|$ denotes the number of users who rate both item i and j, and $|U_i|$ denotes the number of users who rate i. CA is a symmetric function, *i.e.,* $CA(i,j) = CA(j,i)$. $CA \in \mathscr{R}$ and its range is from 0 to 1. When $|U_i|$ and $|U_j|$ are close to $|U_{i,j}|$, CA tends to 1. It means nearly all users who rate i also rate j. By contrast, when CA is small and tends to 0, a few users rate both i and j, *i.e.,* $|U_{i,j}|$ is small. Assume $|U_{i,j}| = |U_{m,n}|$. i but not j is a popular item (*i.e.,* $|U_i| >> |U_j|$), while the popularity of m and n is similar (*i.e.,* $|U_m| \simeq |U_j|$). $CA(m,n)$ is larger than $CA(i,j)$ since a user who rates i is less likely to rate j, which means CA can reduce the influence of the imbalance rating users.

B. Proposed Feature

This section discusses a method to measure the distribution of CA of the rated items of a user profile and all items. Three features are proposed in order to capture the information provided by different arranges of the distribution. Given a user profile database, represented by **R**. A row of **R** represents a profile of a user. $u_i = [r_{i1}, r_{i2}..., r_{im}]$ represent the ith user profile, where $i = 1, 2, ...n$, n is the number of users in **R**, m is the number of items considered in the system, and r_{ij} represents the score of the item j given by the user i. $r_{ij} = \emptyset$ indicates the user i does not rate the item j.

For each profile, u_i is defined as a set U_i containing the items have been rated by the user i, *i.e.,* $U_i = [j | r_{ij} \neq \emptyset, 1 \leq j \leq m]$. CA of each rated item of the user i and all items in the dataset (**R**) of a recommender system is measured. We define $C_i^x = CA(x,i)$, where $i = 1, 2, ...m$. After that, all calculated C_i^x are sorted in ascending orders, such as $C_{s_1}^x, C_{s_2}^x, ..., C_{s_m}^x$ where $C_{s_i}^x \leq C_{s_j}^x$ and $1 \leq i < j \leq m$. In our model, we only consider a particular interval of the distribution of correlation. The interval is defined as a percentage pair (p_l, p_r) containing all real numbers between p_l and p_r. The average of $C_{s_i}^x$ with s_i in the range of $p_l\%$ and $p_r\%$ of the number of all items (m) is shown as follows:

$$S_{(p_l,p_r)}^x = E_{\lfloor p_l \times m \rfloor \leq i \leq \lceil p_r \times m \rceil}(C_{s_i}^x). \tag{2}$$

Three features considering different ranges, including $(0\%, 20\%)$, $(40\%, 60\%)$, and $(80\%, 100\%)$ are defined in (3), (4), and (5). R_L, R_M, and $R_H \in \mathscr{R}$, and their ranges are from 0 to 1. A larger value of R indicates the rated items are highly correlated with other items, correlation with each other, and vice versa.

$$R_L = E_{x \in U_i}(S_{(0\%,20\%)}^x) \tag{3}$$

$$R_M = E_{x \in U_i}(S_{(40\%,60\%)}^x) \tag{4}$$

$$R_H = E_{x \in U_i}(S_{(80\%,100\%)}^x) \tag{5}$$

As rated items are randomly selected in malicious users profiles in many shilling attacks, we expect the distribution of rated item correlation of malicious users should be different from the one of real users. Our proposed features capture three intervals of the distribution, which contains the items with smallest, middle and largest values of CA. The detailed procedure of calculating the proposed features is shown in Algorithm 1. The item correlation matrix based on **R** measured by CA can be calculated offline. And the time complexity of Algorithm 1 is $O(|U_i| \times m)$.

IV. Experiment

The performance of the proposed features is evaluated and compared with the existing features used in shilling attack detection experimentally in this section.

Algorithm 1 Extraction of the proposed features

Input: $CA(i, j)$: the item correlation matrix based on **R**; **R**: a user profile database; U_i: a set containing the items rated by the user i.

Output: R_L, R_M, R_H: the proposed features

```
for each rated item x ∈ U_i do
    for each item i ∈ m do
        C_i^x ← CA(x, i);
    end for
    C_{s_i}^x ← Sort(C_i^x, asc);
    L_x ← E_{⌊0%×m⌋ ≤ i ≤ ⌈20%×m⌉}(C_{s_i}^x);
    M_x ← E_{⌊40%×m⌋ ≤ i ≤ ⌈60%×m⌉}(C_{s_i}^x);
    H_x ← E_{⌊80%×m⌋ ≤ i ≤ ⌈100%×m⌉}(C_{s_i}^x);
end for
R_L ← E(L_x) for x ∈ U_i;
R_M ← E(M_x) for x ∈ U_i;
R_H ← E(H_x) for x ∈ U_i;
Return R_L, R_M, R_H
```

A. Dataset

MovieLens 100K dataset[1] is used in the experiment. It is collected by the GroupLens Research project in the University of Minnesota through the MovieLens website[2] in the seven-month period from September, 1997 to April, 1998. It contains 100,000 rating provided by 943 users on 1,682 movies. Each user rates at least 20 times. Each user can rate a movie from 1 to 5, where a larger value of score represents a deeper degree of favor.

B. Shilling Attack Methods

Random, *average*, *bandwagon* and *AoP 40%* attacks mentioned in Section II-A are considered in our experiment. *Random* and *average* attack select all the items randomly from the whole dataset. For bandwagon attack, 20% of the rated items are randomly selected from the popular item set, and the rest of the filler items are randomly selected from the whole dataset. *AoP 40%* attack first consider the 40% of the most popular items as the selected item set, then select all the rated items randomly from the selected item set.

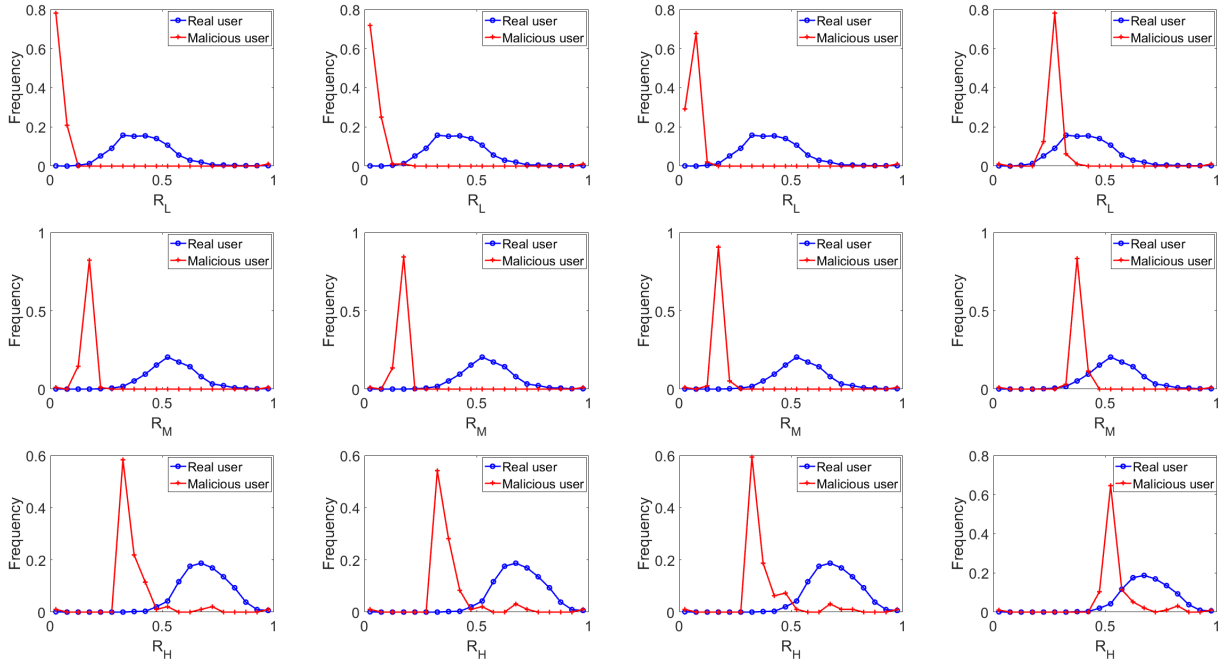
C. Illustration of Discrimination Ability of the Proposed Features

This experiment aims to illustrate the discrimination ability of the proposed features on separating real and malicious user profiles generated by the common shilling attacks.

1) Experimental Setting: All 943 genuine user profiles contained in the dataset are used in this experiment. In addition, malicious profiles with As=10% and Fs ranging from 3% to 20% generated by each of *Random*, *average*, *bandwagon* and *AoP 40%* attacks are injected to the dataset for an experiment.

[1]https://grouplens.org/datasets/movielens/

[2]http://movielens.umn.edu



(a) Random attack (b) Average attack (c) Bandwagon attack (d) AoP 40% attack

Fig. 1. Real and malicious user profiles generated by different shilling attack in the extracted feature space of our proposed features

2) Results and Discussion: The discriminant ability of each proposed feature is investigated individually. The histogram with the interval = 0.05 is applied to represent the distribution of the real (*i.e.,* a blue line) and malicious (*i.e.,* a red line) users generated by different attack methods in terms of each of R_L, R_M, and R_H. The figures are shown in Figure 1.

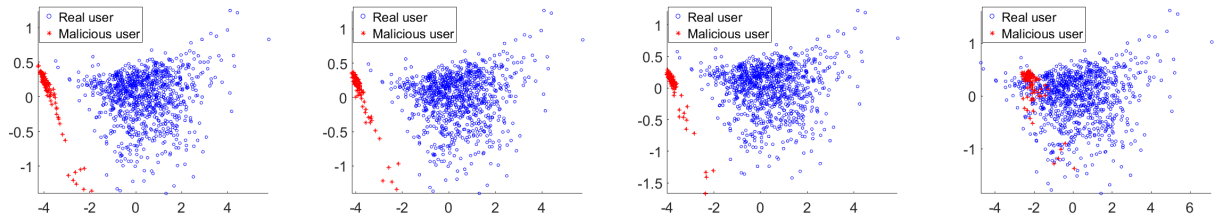
In general, both distributions of real and malicious users generated by the attacks follow the normal distribution in the space of R_L, R_M, and R_H. The mean and standard deviation of distribution of malicious users are smaller than the ones of real users respectively. As genuine users rate items based on their preferences, the rated items have higher correlation with others than the one of malicious users who select rated items randomly. The random selection reduces the rated item correlation in malicious profiles. The weaker discrimination ability of R_H is expected in comparison of R_L and R_M since two types of user profiles overlap each other for each attack. On the other hand, R_L and R_M has the same discrimination ability in both *random* and *average* attack. R_L performs the best among all features (*i.e.,* no overlapping) in *bandwagon* attack, while R_M has the smallest overlapping area in *AoP 40%* attack. The means of distributions of *AoP 40%* attack are relatively larger in a comparison of the other shilling attack methods in all features. It is because *AoP 40%* attack chooses a large amount of popular items which highly correlated with other items. This may increase the difficulty of its detection.

In order to consider the discrimination ability of all features, two features are extracted from the proposed features using Principal Component Analysis (PCA). The malicious (*i.e.,* a red point) and real (*i.e.,* a blue point) user profiles are illustrated in the features extracted from R_L, R_M, and R_H is shown in Figure 2. *Random* and *average* attacks have similar distributions in the feature space because they both select rated items randomly. The classes of malicious profiles of *random*, *average* and *bandwagon* attacks are clearly separated from the class of real users. However, the samples of *AoP 40%* attack overlap with the real user profiles. This observation is consistent with the previous one. Therefore, the experimental results in this section suggest that the proposed features have good discrimination ability in identifying *random*, *average* and *bandwagon* attacks. Although *AoP 40%* attack cannot be clearly separated from the real user profiles, the misclassified region should be small as the malicious user profiles condense as a cluster.

D. Detection Accuracy Evaluation

The performance of supervised learning based detection method with and without using our proposed features is evaluated experimentally in this section.

1) Experimental Setting: SVM is used as a classifier in our experiments. The representative features are selected from each category: DegSim for *profile similarity* base features; RDMA, WDMA, and WDA for *rating behavior* base features; and FSPII and FASUII for *item distribution* base features. The popular (unpopular) item set is defined as the set containing the most 40% popular (unpopular) items for FSPII (FSUII). We consider the model using only these six features, named as



(a) Random attack (b) Average attack (c) Bandwagon attack (d) AoP 40% attack

Fig. 2. Real and malicious user profiles generated by different shilling attack in the extracted feature space of our proposed features

base, the one using our proposed features and the six features, named as *base+R*, and the ones using each of our features and the six features (*base+*R_L, *base+*R_M, and *base+*R_H).

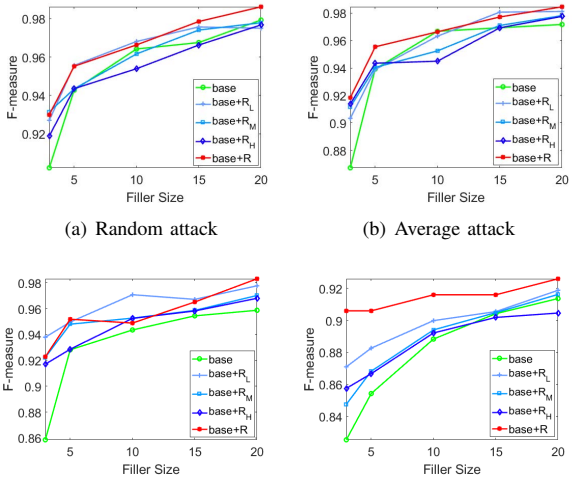
For our experiment, 200 out of 943 user profiles are randomly selected as the training set for the supervised shilling attack detection method. The rest of the profiles is used as the test set. We assume the defender does not obtain complete knowledge on the attack, but only know the shilling attack should be in either *random*, *average*, *bandwagon* and *AoP* 40% attacks. Malicious profiles are generated randomly from the attacks. As and Fs are set as 200, and 3% to 20% of all items in training. As is set as 10% of the test set containing 743 genuine user profiles and a number of filler sizes are used, *i.e.,* $Fs = \{3\%, 5\%, 10\%, 15\%, 20\%\}$, for each attack. For malicious profiles in the training and test sets, the target item is randomly selected to ensure the generality of our results.

The evaluation of the detection performance is *precision*, *recall*, and *F-measure*. F-measure [15] shown in (6) is used due to the imbalance problem of shilling attack detection. The value of F-measure is small either precision and recall is low. Higher F-measure indicates accurate prediction on both classes. All results are obtained by calculating the average values of the experiments repeated five times individually.

$$F\text{-}measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{6}$$

2) Results and Discussion: The detection performance without (*i.e., base*) and with the proposed features (*i.e., base+*R_L, *base+*R_M, *base+*R_H, and *base+R*) are evaluated and compared in this section. The F-measure of these detection methods on four attack models are shown in Figure 3. Y-axis represents the value of F-measure, while X-axis represents the Fs.

In general, the detection accuracy of SVMs with our features is higher than the one without using the features in terms of F-measure, especially when the filler size is small. For example, when $Fs = 0.03$, F-measure increases more than 6% by using our features. Although in some cases, *e.g.,* when $Fs = 0.10$ in *average* attack, *base* is more accurate than others, the general performance of *base+*R_L, *base+*R_M, *base+*R_H, and *base+R* are better. *base+*R_L have the highest F-measure in comparison with *base+*R_M and *base+*R_H. It confirms the observation in the last experiment which shows the overlapping area between the distributions of real and malicious user profiles is smallest among all features. For all methods, F-measure increase with the increases of Fs since more information is provided by a malicious with more rated items. In addition, *base+R* achieves the best performance among all methods in average. As F-measure of *base+R* is larger than *base* under the attacks with all values of Fs, which confirms that the proposed three features are able to improve the detection performance.



(a) Random attack (b) Average attack

(c) Bandwagon attack (d) AoP 40% attack

Fig. 3. F-measure of *base*, *base+*R_L, *base+*R_M, *base+*R_H, and *base+R* on different shilling attacks

Precision and recall of *base+R* and *base* are given in Table I for detailed comparison. The bold value indicates the higher precision or recall value in comparison of *base+R* and *base* in the same attack setting. The value with $*$ indicates the difference between the two methods is statistically significant with 95% confidence using Student's t-test. In general, the performance of both methods in terms of precision and recall is consistent with the ones evaluated by F-measure. Our proposed features successfully improves the detection accuracy on shilling attack in most cases, especially in the case with smaller Fs. The accuracy of detection on *bandwagon* and *AoP 40%* attack is lower than the one of *random* and *average* attack. As a result, the experimental results suggest that our proposed features increase the accuracy of detection on shilling attack efficiently.

TABLE I
PRECISION AND RECALL OF *base* AND *base+R* IN DIFFERENT SETTINGS

Filler size			0.03	0.05	0.10	0.15	0.20
Random	Precision	base	91.1	92.6	93.1	93.9	95.4
		base+R	**92.4***	92.6	93.1	**95.6***	**96.9***
	Recall	base	89.3	96.0	100	100	100
		base+R	**93.1***	**98.4***	100	100	98.4
Average	Precision	base	88.7	92.2	93.6	94.0	94.5
		base+R	**91.5***	92.2	**93.1**	**95.0***	**97.0***
	Recall	base	84.8	95.7	100	100	100
		base+R	**92.8***	**99.2***	100	100	100
Bandwagon	Precision	base	90.3	91.3	93.0	92.1	91.3
		base+R	**91.5***	**92.7***	**93.1**	**95.2***	**97.6***
	Recall	base	81.9	94.4	100	100	100
		base+R	**94.2***	**98.3***	100	100	99.7
AoP 40%	Precision	base	71.4	74.6	79.9	82.5	85.1
		base+R	**84.3***	**82.8***	**84.5***	**84.9***	**86.2***
	Recall	base	97.7	100	100	100	98.7
		base+R	**97.8**	100	100	100	**100**

* Statistically significant difference with 95% confidence in comparison of *base* and *base+R* using the Student's t-test.

V. CONCLUSION

Shilling attack is one of the major threats to CF recommender systems. By injecting a small number of malicious profiles, recommendation results may be influenced significantly. Shilling attack significantly reduces the trustworthiness of a recommender system. Therefore, there is an emergent need for a well crafted shilling attack detection method in order to decease the impact of the attack. The features proposed for supervised learning based detection method in previous studies measure the characteristic of rating values or items of a profile individually, and relation among items is neglected. However, we observe that real users rate an item based on their preferences but most existing attack methods select rated item randomly. Their correlation of rated items should be different. In this study, we proposed three features to capture the rated item correlation in terms of CA. We firstly illustrate the discrimination ability of the proposed features, and evaluate their performance by using MovieLens 100K dataset. The experimental results confirm that our proposed features have significant contribution to the shilling attack detection. One possible feature work is to measure additional information from the distribution of rated item correlation in order to improve the generalization ability of shilling attack detection.

VI. ACKNOWLEDGEMENT

This paper is supported by the Natural Science Foundation of Guangdong Province, China (No. 2018A030313203).

REFERENCES

[1] Desrosiers, Christian, and G. Karypis. A Comprehensive Survey of Neighborhood-based Recommendation Methods. *Recommender Systems Handbook* Springer US, (2011):107-144.

[2] Chen K, Patrick P. K. Chan, et al. Shilling Attack Based on Item Popularity and Rated Item Correlation against Collaborative Filtering. *International Journal of Machine Learning and Cybernetics.* in press

[3] Gunes, Ihsan, et al. Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review* 42.4(2014):767-799.

[4] Mehta, Bhaskar, and W. Nejdl. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User-Adapted Interaction* 19.1-2(2009):65-97.

[5] Zhou W, Wen J, Xiong Q, et al. SVM-TIA a shilling attack detection method based on SVM and target item analysis in recommender systems. *Neurocomputing* 210.C(2016):197-205.

[6] Chirita, Paul Alexandru, W. Nejdl, and C. Zamfir. Preventing shilling attacks in online recommender systems. (2005):67-74.

[7] Burke, Robin, et al. Classification features for attack detection in collaborative recommender systems. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM,* (2006):542-547.

[8] Zhang, Fuzhi, and Q. Zhou. HHTCSVM: An online method for detecting profile injection attacks in collaborative recommender systems. *Knowledge-Based Systems* 65.4(2014):96-105.

[9] Agrawal, Rakesh, T. Imielinski, and A. N. Swami. Mining Association Rules Between Sets of Items in Large Databases, *SIGMOD Conference. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data.* (1993):207–216.

[10] Lam, Shyong K., and J. Riedl. Shilling recommender systems for fun and profit. *International Conference on World Wide Web ACM,* (2004):393-402.

[11] Seminario, Carlos E., and D. C. Wilson. Attacking item-based recommender systems with power items. *ACM Conference on Recommender Systems ACM,* (2014):57-64.

[12] Mehta, Bhaskar, and W. Nejdl. Unsupervised strategies for shilling detection and robust collaborative filtering. *User Modeling and User-Adapted Interaction* 19.1-2(2009):65-97.

[13] Hurley, Neil, Z. Cheng, and M. Zhang. Statistical attack detection. *ACM Conference on Recommender Systems, Recsys 2009, New York, Ny, Usa, October DBLP,* (2009):149-156.

[14] Yang, Zhihai, and Z. Cai. Detecting abnormal profiles in collaborative filtering recommender systems. *Journal of Intelligent Information Systems* (2015):1-20.

[15] Maratea, Antonio, A. Petrosino, and M. Manzo. Adjusted F-measure and kernel scaling for imbalanced data learning. *Information Sciences* 257.2(2014):331-341.