A State of Art Survey on Shilling Attack in Collaborative Filtering Based Recommendation System

Krupa Patel, Amit Thakkar, Chandni Shah and Kamlesh Makvana

Abstract Recommendation system is a special type of information filtering system that attempts to present information/objects that are likely to the interest of user. Any organization, provides correct recommendation is necessary for maintain the trust of their customers. Collaborative filtering based algorithms are most widely used algorithms for recommendation system. However, recommender systems supported collaborative filtering are known to be extremely prone to attacks. Attackers will insert biased profile information or fake profile to have a big impact on the recommendations made. This paper provide survey on effect of shilling attack in recommendation systems, types of attack, knowledge required and existing shilling attack detection methods.

Keywords Recommendation system · Collaborative filtering shilling attack · Detection and evaluation parameters · Information filtering

1 Introduction

Recommendation systems (RS) provide information or item that is interest of the user by analysing rating pattern and stable information of user. The huge growths of information on the web as well as variety of guests to websites add some key challenges to recommender systems technology; these are producing accurate recommendation and handling several recommendations with efficiency [1].

K. Patel (✉) · A. Thakkar · C. Shah · K. Makvana
Department of Information Technology, CSPIT, CHARUSAT, Anand, India
e-mail: 14pgit010@charusat.edu.in

A. Thakkar
e-mail: amitthakkar.it@charusat.ac.in

C. Shah
e-mail: chandnishah.it@charusat.ac.in

K. Makvana
e-mail: kamleshmakvana.it@charusat.ac.in

S.C. Satapathy and S. Das (eds.), *Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 1*, Smart Innovation, Systems and Technologies 50, DOI 10.1007/978-3-319-30933-0_38

Therefore, new recommender system technologies are required which will quickly turn out prime quality recommendations even for immense information sets.

Content based and collaborative filtering (CF) based are two approaches for developing recommendation systems. In content based system items are recommended based on users past rating history and content of items. Collaborative filtering recommendation system is based on U-I rating matrix. In a typical Collaborative filtering system, an $n \times m$ user-item matrix is created, where n users' preferences about m products are represented as ratings, either numeric or binary. To obtain a prediction for a target item i or a sorted list of items that might be liked, an active user u sends her known ratings and a query to the system. CF system estimates similarities between u and each user in the database, forms a neighbourhood by selecting the best similar users, and estimate a prediction (p_{ui}) or a recommendation list (top-N recommendation) using a CF algorithm [2, 3]. Profile injection attacks degrade the quality and accuracy of a CF based recommender system over inflicting frustration for its users and probably resulting in high user defection [4]. CF based recommendation systems are extremely prone for shilling attacks then content based recommendation systems [2]. New technology are needed that cannot biased to the various fake profile, and generate recommendation with high precision.

Overall success of CF based recommendation system is depends on how it handle shilling attack and how effectively detect shilling attacks [5]. In this paper we provide survey of various types of shilling attack in CF based RS. Also classification of shilling attacks, detection attributes detection techniques and some evaluation parameters of recommendation systems.

The paper is designed, as follows: In Sect. 2 we briefly discuss theoretical background after that in Sect. 3, contain related work then in Sect. 4 contain various detection attributes of shilling attack detection after that in Sect. 5 evaluation matrix and parameters and then in Sect. 6 we discuss some future work and open issues. Finally, we conclude our paper in Sect. 7.

2 Theoretical Background

Section 1, provides basic introduction about recommendation system so now, we focus on shilling attack in collaborative filtering based recommendation system.

2.1 Shilling Attacks

Recommendation schemes are successful in e-commerce sites; they are prone to shilling or profile injection attacks. Shilling attack or profile injection attacks is outlined as,

Table 1 U-I matrix without Shilling attack

User	Items						Similarity with user 1
	I1	I2	I3	I4	I5	I6	
U1	5	2	3	3		?	1.00
U2	2		4		4	1	−1.00
U3	3	1	3		1	2	0.76
U4	4	2	3	1		1	0.72
U5	4	3		3	3	2	0.94

Table 2 U-I matrix with Shilling attack

User	Items						Similarity with user 1
	I1	I2	I3	I4	I5	I6	
U1	5	2	3	3		?	1.00
U2	2		4		4	1	−1.00
U3	3	1	3		1	2	0.76
U4	4	2	3	1		1	0.72
U5	4	3		3	3	2	0.94
U6	4	2		3	3	5	0.98

Malicious users and/or competitive vendors may attempt to insert fake profiles into the user-item matrix in such a way so they will have an effect on the predicted ratings on behalf of their benefits [2].

To understand how shilling attack works, consider Table 1. Contain 6 users and 6 items and we want to predict rating on item 6 by user 1 which is our target user.

Without shilling attack similarity with user 1 is given in Table 1. This similarity is calculated using Pearson correlation coefficient (PCC). If we chose k = 1 then most similar user with user 1 is user 5 and rating by user 1 on item 5 is 2, which is our correct answer.

Now, if attacker enters shilling attack profile which is user 6 then similarity with target user 1 is shown in Table 2. Here with shilling attack profile user 6 is most similar user with target user 1 with similarity 0.98, and rating for item 6 by user 1 is 5 instead of 2 (original rating without shilling attack).

Hence, shilling attacks is reducing quality of data and hence reduce accuracy of recommendation system.

2.2 Classification of Shilling Attacks

Shilling attacks are classified based on intent and based on amount of knowledge required to build shilling attack profiles.

Based on intent. Based on intent shilling attacks are classified as push and nuke attacks. Push attack try to increase popularity of target item by giving high rating to

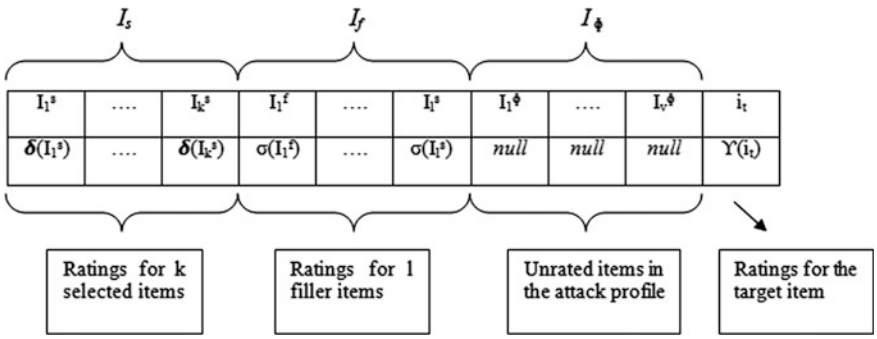

Fig. 1 Shilling attack profile

target item. Nuke attack tries to reduce popularity of target item by giving low rating to target item [6].

Based on knowledge required. Based on amount of knowledge required there are different shilling attack models like *average attack, random attack, bandwagon attack, reverse bandwagon, segment attack* etc.

Attack profile for shilling attack is shown in Fig. 1.

The attack profile is an m-dimensional vector of ratings as per Fig. 1, were m is the total range of items within the system. The profile is partitioned into four components. The empty partition, I_ϕ is those items with no ratings in the profile. The only target item i_t will be given a rating as determined by the function Υ, usually this will be either the highest or lowest possible rating, looking on the attack type (push/nuke). As described below, some attacks need distinguishing a group of items for special care during the attack. This special set I_s receives ratings as given by the function δ. Finally, there is a group of filler items I_f whose ratings are given as specified by the function σ. It is the strategy for choosing items in I_s and I_f and the functions Υ, σ, and δ that outline an attack model and provides it its character [6].

For different attack models different strategies are used for creating attacker profiles and how to provide rating for items to create attacker profiles are shown in Table 3.

Table 3 Attack profile summary [2]

Attack model	I_s		I_f		I_ϕ	I_t (push/nuke)
	Items	Ratings	Items	Ratings		
Random	Null	–	Randomly chosen	System mean	Null	r_{max}/r_{min}
Average	Null	–	Randomly chosen	Item mean	Null	r_{max}/r_{min}
Bandwagon	Popular items	r_{max}/r_{min}	Randomly chosen	System mean/item mean	Null	r_{max}/r_{min}
Reverse bandwagon	Unpopular items	r_{max}/r_{min}	Randomly chosen	System mean	Null	r_{max}/r_{min}
Segment	Segmented items	r_{max}/r_{min}	Randomly chosen	r_{max}/r_{min}	Null	r_{max}/r_{min}

3 Related Work

In this section we have a tendency to represent some related works in field of shilling attacks in recommendation systems. Since shilling profiles look like authentic profiles, it is very tough to spot them. To discover shilling profile various statistical, classification, clustering techniques are used. Bryan et al. [7] Suggest new algorithm known as "Unsupervised Retrieval of Attack Profiles" (UnRAP). They recommend new measure known as Hv-score measure to find shilling profile from genuine profile. They said that Hv-score value of attacker profile is higher than genuine profile. Based on this assumption they identify attacker profile. Lu [8] Extends work of [7] to find group of attacker instead of individual attackers. With help of various detection matrices and analysing raring pattern of attacker [9] Propose unsupervised learning method for detection of fake profile using target item analysis. Algorithm find potential attack profiles using digsim and rdma (Rating Deviation from Mean Agreement) and then refine set of potential profile using target item analysis. Supervised learning is another approach to detect shilling attack in memory based CF. Zhang and Zhou [10] suggest Ensemble learning concept for shilling attack detection using back propagation neural network classifier, finally output is combined using voting strategy. Semi-supervised learning also helpful to detect shilling profiles. Bilge et al. [11] Use bisecting k-means algorithm to generate binary decision tree. Intra cluster correlation (ICC) is used to find correlation within cluster between the profiles. This method assumes that attacker profiles in cluster have high ICC between them. And cluster with high value for ICC is considered as attacker cluster. But Performance of this scheme is slightly worse with increasing filler size in segment attack. Zhang et al. [12] Detect shilling attacks using clustering social trust information between the users. They propose two algorithms, CluTr and WCluTr, to mix clustering with "trust" among users. According to them user with no incoming trust is considered as attacker profiles. Cao et al. [13] Use Semi-supervised learning method semi-SAD. Combination of EM-λ and naïve-bayes is used for detection of shilling attacks.

4 Detection of Shilling Attack

CF based recommendation systems are vulnerable to shilling attacks. We begin this section with a review of some of the statistical measures that have been designed to detect shilling attack in recommendation system. Some of the Standard shilling attack detection metrics are explains below:

Rating Deviation from Mean Agreement (*RDMA*). This measures a user's rating disagreement with other genuine users in the system, weighted by the inverse number of item that user rated. It is defined as,

$$RDMA_u = \frac{\sum_{i=0}^{N_u} \frac{|r_{u,i} - Avg_i|}{NR_i}}{N_u} \quad (1)$$

Weighted Deviation from Mean Agreement (*WDMA*). This measure is strongly based on RDMA; however it places higher weight on rating deviations for sparse items [6]. It is defined as,

$$WDMA_u = \frac{\sum_{i=0}^{N_u} \frac{|r_{u,i} - Avg_i|}{NR_i^2}}{N_u} \quad (2)$$

Where, N_u is the range of items user u rated, $r_{u,i}$ is the rating given by user u to item i, NR_i is the overall range of ratings in the system given to item i. Avg_i is average rating of item i.

Degree of similarity (*DegSim*). Which based on hypothesis that is attacker profiles is highly similar with each other because of theirs characteristics and they are generated with same process [7]. But this profile has low similarity value with genuine profiles. It can be defined as,

$$DigSim_u = \frac{\sum_{v \epsilon\, neighbors(u)} W_{u,v}}{k} \quad (3)$$

Where, $W_{u,v}$ similarity between u and k-nearest neighbours v. and k is number of nearest neighbours of user u.

Equation for similarity between u and v using Pearson correlation coefficient is given as below [9],

$$W_{u,v} = \frac{\sum_{i \epsilon I} \left(r_{u,i} - \overline{r_u}\right)\left(r_{v,i} - \overline{r_v}\right)}{\sqrt{\sum_{i \epsilon I} \left(r_{u,i} - \overline{r_u}\right)^2 \left(r_{v,i} - \overline{r_v}\right)^2}} \quad (4)$$

Where, I is the set of items that users u and v both rated, r_{ui} is the rating user u gave to item i, and $\overline{r_u}$ is the average rating of user u.

Length Variance (*LengthVar*). This attribute relies on the length of user profile. Most of the attacker enters shilling profile that contains large number of rated items [6]. Thus shilling profile has high value for this attribute. Length Variance (LengthVar) that is a measure of what proportion the length of a given profile varies from the average length within the database. It is defined as,

$$LengthVar_u = \frac{|n_u - \overline{n_u}|}{\sum_{u \epsilon U} \left(n_u - \overline{n_u}\right)^2} \quad (5)$$

Where, nu is the average length of a profile in the system.

5 Evaluation Parameters and Matrices

Various evaluation matrices are used for evaluation of effect of shilling attack in recommendation system and evaluating detection algorithms and measure accuracy of recommendation system. These measures are shown in Table 4.

6 Discussion and Open Issues

The internet has been increasing attention now a days. Due to the continuously increasing popularity of internet large amount of data are available. This large data create information overload problem. Recommendation system is system that predict users interest and recommend items to the user. Therefore, the research about recommender systems seems to remain popular. Similarly, shilling attacks against such systems will be in place, as well. Hence, number of researchers are doing research in this area and they are try to find various solutions but there are still missing gaps in this area that is need to be filled.

From all above surveys in Sect. 3 some interesting points are found these are almost all detection methods are unsupervised learning methods. Using supervised learning high accuracy is possible to achieve then unsupervised method. Detection

Table 4 Evaluation parameters [2]

Parameter	Significance	Equation
Precision	Precision (also referred to as positive predictive value) is that the fraction of retrieved instances (attacker) that are really attacker	$precision = \frac{TP}{TP+FP}$
Recall	Recall (also called sensitivity) is that the fraction of relevant instances (attacker) that area unit retrieved as attacker	$recall = \frac{TP}{TP+FN}$
F1 measure	Combination of precision and recall Use for Accuracy of detection algorithm	$F1 = \frac{2*precision*recall}{precision+recall}$
Prediction shift	Prediction shift is that the average change within the predicted rating for the attacked item before and when the attack. This measure is employed for assessing impact of shilling attack	$prediction\ shift = r_{u,i} - r'_{u,i}$ $r_{u,i}$ is rating before shilling attack $r_{u,i}'$ is rating after shilling attack
MAE (mean absolute error)	MAE measures however close the estimated predictions to their discovered ones	$MAE = \frac{1}{N}\sum_{t=1}^{N}\frac{\lvert A_t - F_t\rvert}{A_t}$ A_t = actual value F_t = predicted value

of fake profile in model based recommendation is also one interesting point. Sparse database are vulnerable to shilling attack hence this direction is also need to be investigated. Design a various methods that effectively improves recommendation accuracy in presence of sparsity and shilling attack profile are one of the good idea of research. Using social relationship between users we can also find fake profile that help to improve recommendation accuracy. Using content and social information of user analyze their search history to determine that given user is attacker or not this is also one of the new topic of research.

7 Conclusion and Future Work

In this survey we discuss about effect of shilling attack in recommendation system, their types, detection parameters, evaluation parameters and related works that was done in this In future we are planning to conduct detail survey in field of fake review detection in model based and hybrid recommendation system. We are also planning to propose method for detection of shilling attack using supervised learning.

References

1. Almazro, D., Shahatah, G., Albdulkarim, L., Kherees, M., Martinez, R., Nzoukou, W.: A Survey Paper on Recommender Systems (2010)
2. Gunes, I., Kaleli, C., Bilge, A., Polat, H.: Shilling attacks against recommender systems: a comprehensive survey. Artif. Intell. Rev., 1–33 (2012)
3. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
4. Sandvig, J., Mobasher, B., Burke, R.: A survey of collaborative recommendation and the robustness of model-based algorithms. IEEE Data Eng. Bull., 1–11 (2008)
5. Asanov, D.: Algorithms and Methods in Recommender Systems (2011)
6. Burke, R., Mobasher, B. Williams, C., Bhaumik, R.: Classification features for attack detection in collaborative recommender systems. In: Proceedings of the 12th ACM SIGKDD International Conference Knowledge Discovry and Data Mining KDD 06, p. 542 (2006)
7. Bryan, K., O'Mahony, M., Cunningham, P.: Unsupervised retrieval of attack profiles in collaborative recommender systems. In: Recsys'08 Proceedings of the 2008 ACM Conference on Recommender Systems, pp. 155–162 (2008)
8. Lu, G.: Engineering, "a group attack detector for collaborative filtering recommendation." IEEE Internet Comput., 2–5 (2014)
9. Zhou, W., Wen, J., Koh, Y.S., Xiong, Q., Gao, M., Dobbie, G., Alam, S.: Shilling attacks detection in recommender systems based on target item analysis. PLoS ONE **10**(7), e0130968 (2015)
10. Zhang, F., Zhou, Q.: Ensemble Detection Model for Profile Injection Attacks in Collaborative Recommender Systems Based on BP Neural Network, vol. 9, pp. 24–31 (2015)

11. Bilge, A., Ozdemir, Z., Polat, H.: A novel shilling attack detection method, Procedia Comput. Sci. **31**, 165–174 (2014)
12. Zhang, X.L., Lee, T.M.D., Pitsilis, G.: Securing recommender systems against shilling attacks using social-based clustering. J. Comput. Sci. Technol. **28**, 616–624 (2013)
13. Cao, J., Wu, Z., Mao, B., Zhang, Y.: Shilling attack detection utilizing semi-supervised learning method for collaborative recommender system. World Wide Web **16**, 729–748 (2013)