International Journal of Hybrid Information Technology
Vol.9, No.2 (2016), pp. 381-388
http://dx.doi.org/10.14257/ijhit.2016.9.2.34

Segment-Focused Shilling Attacks against Recommendation Algorithms in Binary Ratings-based Recommender Systems

Fuguo Zhang[1,2]

[1](*School of Information Technology, Jiangxi University of Finance & Economics, Nanchang, China)*
[2](*Jiangxi Key Laboratory of Data and Knowledge Engineering, Jiangxi University of Finance and Economics, Nanchang, China*)
Email: redbird_mail@163.com

Abstract

As one of the most successful approaches to building recommender systems, collaborative filtering (CF) uses the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users. However, the open nature of collaborative filtering recommender systems allows attackers who inject biased profile data to have a significant impact on the recommendations produced. Shilling attacks against numeric ratings-based CF schemes have been extensively studied. To the best of our knowledge, there is rare study about how to attack binary ratings-based recommendation systems. Hence, shilling attacks strategies against binary ratings-based recommendation algorithms need to be further investigated. The empirical results obtained from Movielens dataset show that the segment attack, which is easy to mount, affects strongly against BU-CF algorithm and BI-CF algorithm, and the attack size and the filler size don't have the same sensitivity in attack effect.

Keywords*: collaborative filtering, segment attack model, binary ratings, recommender system*

1. Introduction

Many electronic commerce sites offer a large amount of various products for sale. Choosing among so many options is challenging for consumers. Recommender systems have emerged in response to this problem, and become an integral part of some e-commerce sites [1]. They receive information from a consumer about which products she is interested in and help customers find which products they would like to purchase at E-Commerce sites. Today, recommender systems are deployed on hundreds of different sites, serving millions of consumers. Many electronic commerce sites already benefit from novel opportunities of personalized marketing leverage offered by these information systems [2].

A variety of techniques have been proposed for performing recommendation, including content-based [3-6], collaborative filtering [7-10], and hybrid recommendation techniques [11-14]. As one of the most successful approaches to building recommender systems, collaborative filtering (CF) is probably the most familiar, most widely implemented and most mature of the technologies [15, 16]. Collaborative filtering uses the user rating data to calculate the similarity or weight between users or items and make predictions or recommendations according to those calculated similarity values. A typical user profile in a collaborative system consists of a vector of items and their ratings. Collaborative recommender systems aggregate ratings or recommendations of objects, recognize commonalities between users on the basis of their ratings, and generate new recommendations based on inter-user comparisons. However, while recommender

ISSN: 1738-9968 IJHIT

systems are helping the customers find things that they might want to buy, they can also be a valuable asset for retail companies to make more money by selling more products to customers, accordingly, producers of items have a natural motivation to have their products recommended more often than those of their competitors [17]. Because of the open nature of collaborative filtering algorithms, unscrupulous producers can easily introduce biased data in an attempt to force the system to "adapt" in a manner advantageous to them [18].

Users' preferences might be represented using either numeric or binary ratings. Numeric rating scales are widely used online in an attempt to provide indications of consumer opinions of products. For example, big e-commerce sites like Amazon and eBay utilize the five-star scale, and Netflix also uses a five-star scale to power its review system. Most research focus attention on examining the robustness of numeric ratings-based collaborative filtering algorithms in the face of profile injected attacks [14,18-20], furthermore, many detection methods [21-24] in such numeric systems have been proposed. In some cases, ratings may be binary (like/dislike). For instance, YouTube now operates on a thumbs up or down rating scale. It is much easier for a person to declare. However, as far as we know, there is rare study [14] discussing how to attack binary ratings-based recommendation systems. Long and Hu [25] examined the robustness of collaborative filtering algorithms against random attack, bandwagon attack and average attack in a numeric ratings recommender system by calculating the similarity based on binary ratings, but the recommendation prediction algorithms were still based on numeric ratings. The segment attack model is to make inserted bots more similar to the segment market users, and it is easy to mount [26, 27]. Hence, in this paper, we focus on examining the robustness of segment attack against the binary ratings-based collaborative filtering algorithms including BU-CF algorithm and BI-CF algorithm.

Remaining parts of this paper are organized as follows. Section 2 introduces BU-CF and BI-CF algorithms. The segment attack model is described in Section 3. Section 4 presents our experimental work including details of our data set, evaluation metrics, results of different experiments. Finally, in Section 5 we provide some conclusions and future work.

2. Binary Ratings based CF Algorithms

A binary ratings-based system can be represented by a bipartite network G(U,O,E), where U ={u1,u2, . . . ,un}, O = {o1,o2, . . . ,om}and E ={e1, e2, …,el} are the sets of users, objects and links. It can be fully described by an adjacency matrix A ={ $a_{i\alpha}$ }m,n, where the element $a_{i\alpha}$ =1 if a user i has collected an item α, and $a_{i\alpha}$ = 0 otherwise. The most well-known CF algorithms are nearest neighbor algorithms. The two different classes of nearest neighbor CF algorithms: user-based nearest neighbor and item-based nearest neighbor are often discussed in the previous study of shilling attack against recommender systems. There exist some differences between the nearest neighbor CF algorithms in a binary ratings-based system and that in a numeric ratings based system.

2.1. Binary Ratings-based U-CF(BU-CF) Algorithm

User-based algorithms generate a prediction for an item i by analyzing ratings for i from users in u's neighborhood. There are many methods to calculate the similarity. It is often computed using Pearson's correlation coefficient, but it is not fit for a binary ratings-based system. In this paper, we selected the cosine similarity [28] between the

$$s_{ij} = \frac{|\Gamma_i \cap \Gamma_j|}{\sqrt{k_i . k_j}} \quad (1)$$

target user i and a neighbor j, sij, defined as:

where Γ_i and Γ_j denotes the set of objects that user i and user j has collected, respectively; and ki = $|\Gamma_i|$ and kj = $|\Gamma_j|$ are the degree of user i and user j, respectively; namely the number of objects that user i and user j have collected, respectively.

In BU-CF, the predicted score $P_{i\alpha}$ is given as:

$$P_{i\alpha} = \sum_{j=1,\, j\neq i}^{m} s_{ij} a_{j\alpha} \tag{2}$$

where sij denotes the similarity between user i and user j.

2.2 Binary Ratings-based I-CF(BI-CF) Algorithm

For item-based prediction, by using the cosine expression, the similarity between two objects, oi and oj , can be written as:

$$s_{\alpha\beta} = \frac{|\Gamma_\alpha \cap \Gamma_\beta|}{\sqrt{k_\alpha . k_\beta}} \tag{3}$$

where Γ_α and Γ_β denotes the set of the users who has collected object α and object β, respectively; and ki = $|\Gamma_i|$ and kj = $|\Gamma_j|$ are the degree of object i and object j, respectively; similar to the BU-CF algorithm, the recommendation list in a binary ratings-based system can also be obtained by item-based collaborative filtering (BI-CF) algorithm. That is to say,

$$P_{i\alpha} = \sum_{\beta=1,\beta\neq\alpha}^{n} s_{\alpha\beta} a_{i\beta} \tag{4}$$

the user will be recommended objects similar to the ones he/she preferred in the past [29].

3. Segment Attack Model

Item	I_1	…	I_l	I_{1+1}	…	I_k	I_{k+1}	...	I_m
Rating	*1*	…	*1*	null	…	null	*1*	*1*	*1*
		Filler items			Unrated items			Target items	

Figure 1. The General form of a Segment Attack Profile in a Binary

An attack model is an approach to constructing attack profiles based on knowledge of the recommender system, its rating database, its products, and/or its users [30]. Shilling attacks can be classified as push and nuke attacks according to their intent [20]. Push attacks try to make one or more target items recommended to more users, while nuke attacks try to cause them less likely to be recommended. An attack against a collaborative filtering recommender system consists of a set of attack profiles. An attack profile consists of an m-dimensional vector of ratings, where m is the total number of items in the system. The general form of an attack profile is shown in Figure 1. The profile is partitioned in three parts. Firstly, the unrated items partition are those items with no ratings in the profile. Secondly, the target items will be given a rating designed to bias its recommendations, generally they will be either the maximum (push attack) or minimum (nuke attack) rating depending on the attack type. Finally, the set of filler items represent a group of selected items in the database which are assigned ratings within the attack profile, they will be given a rating according to the attack strategy [20].

There has been a number of attack types studied in previous work. From the perspective of an attacker, Segment attack model is more likely that he wish to promote the target product will be interested not in how often it is recommended to all users, but

how often it is recommended to the particular market segment that is likely to already have a propensity to purchase it. For example, Suppose attacker A is a seller of children books. B is a buyer of Harry Potter books, while C is a buyer of books about program design or car repair. There is no doubt that attacker A wish his books be recommended to B rather than C. Therefore, different from the other styles of attack model, the segment attack shifts the focus from trying to impact the complete user set to instead targeting a segment of profiles with specific interests [26, 27]. Meanwhile in this case the attacker does not need to know anything system-specific, merely that certain items are the popular ones that users are likely to have rated. The attacker selects k such items that will be rated highly along with the target item.

4. Experiments

4.1. Dataset and Metrics

In our experiments we have used the publicly -available MovieLens 100K dataset. This dataset consists of 100,000 ratings on 1682 movies by 943 users. Ratings are discrete-valued between 1 and 5. In order to transform the origin numeric ratings-based dataset to a binary ratings-based system, coarse-graining method was applied: a movie has been collected by a user if the giving rating is more than 3.

To measure how effective an attack is in accomplishing its goal, prediction shift [13] and hit ratio [14] are two common metrics to evaluate in change in performance induced by the attack instead of raw performance. However, prediction shift represents the difference in the system's predicted rating for an item before and after the attack. In a binary system, predicted value is a relative value instead of rating.

The hit ratio metric denotes the average likelihood that a top N recommender will recommend the pushed item. It is defined as the sum of the occurrence number of all target items in a top-N recommendation list across all test users divided by the number of pushed items and the number of test users. Suppose Ru be the set of top N recommendations for user u. The hit ratio for a pushed item i over all the test users,

$$HitRatio_i = \sum_{u \in U_t} H_{ui} / | U_t | \quad (5)$$

$HitRatio_i$, can then be computed as:

where H_{ui} denotes the value of a recommendation hit on item i for user u. H_{ui} =1 if $i \in R_u$, and Hui = 0 otherwise. Ut denotes the set of the test users. The hit ratio of the recommender system is defined as follows:
where It is the set of the target items.

4.2. Attack Experimental Design

$$HitRatio = \sum_{i \in I_t} HitRatio_i / | I_t | \quad (6)$$

This paper aims to empirically analyze the effectiveness of the segment attack against BU-CF and BI-CF algorithms. We divided the full dataset into training and test sets. For the segment attack, we randomly selected 50 of the users who has collected any three of the five most popular horror movies, namely, Alien, Psycho, The Shining, Jaws, and The Birds as our targeted market segment user set, and manually selected 30 items from the horror movies which degree is less than 100 for our target items set. The edges between the attack users and the items in the target item set and horror popular item set are connected. The remainder of user profiles after removing the test set is designated as the training set. For all the attacks, we inserted a number of attack profiles into the training set and then generated predictions. To evaluate the sensitivity of attack size and top-N value, we have tested 5%, 10%, 15%, 20%, 25% , 30% attack size and 10, 20, 30, 40, 50

top-N value on each attack type. In addition, to evaluate the sensitivity of filler size we set the filler size from 5 to 30 step 2.

4.3. Results and Discussion

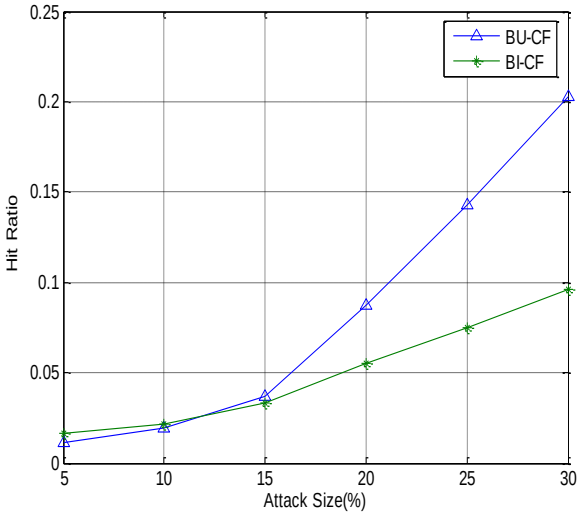

Figure 2. Comparison of the Hit Ratio of the Segment Attack Against BU-CF Algorithm and BI-CF Algorithm at Different Attack Size (Filler Size=15, top n=30)

In our first set of experiments, we compare the hit ratio of the segment attack against BU-CF and BI-CF algorithms at different attack sizes. The results are shown in Figure 2 (push attack). Clearly, the segment attack affected the hit ratio severely. When the attack size is more than 15%, BU-CF algorithm reflects very more strongly than BI-CF algorithm.

The next aspect we examine is the impact on hit ratio with different top-N value. Figure 3 presents hit ratios in different top-N at 15% attack size and 15 popular segment filler size. We can also find that the segment attack is very more effective against BU-CF algorithm than BI-CF algorithm, especially when topN is bigger than 20.

Finally, the comparison results of the hit ratio of the segment attack against BU-CF algorithm and BI-CF algorithm at different topN(Attack Size=15%, Filler Size=15) are shown in Figure 4. For BU-CF algorithm, the hit ratio increases fast during the filler size is smaller than 15, but it changes little when the filler size is bigger than 15. Similarly, the hit ratio of segment attack against BI-CF algorithm reaches the peak when the filler size is 27.

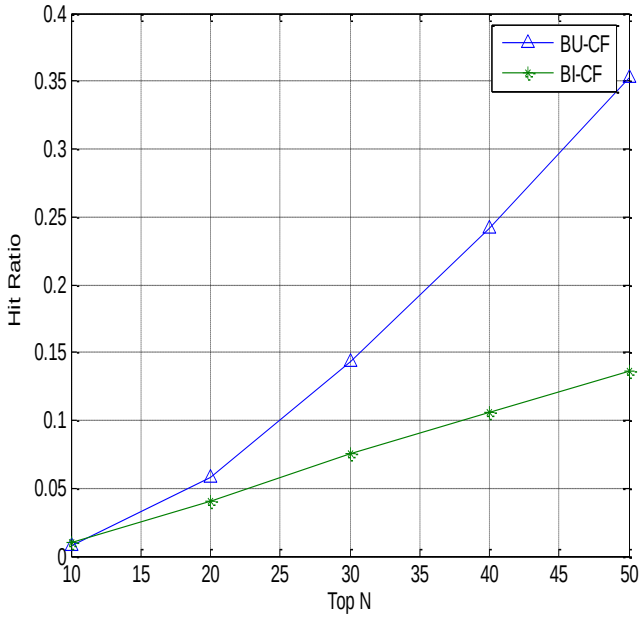

Figure 3. Comparison of the Hit Ratio of the Segment Attack Against BU-CF Algorithm and BI-CF Algorithm at Different topn (Attack Size=15%, Filler Size=15)

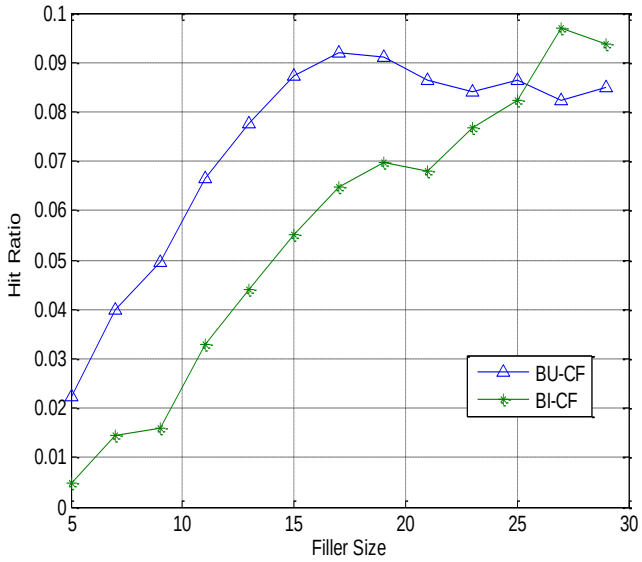

Figure 4. Comparison of the Hit Ratio of the Segment Attack Against BU-CF Algorithm and BI-CF Algorithm at Different Filler Size (Attack Size=20%, top n=30)

5. Conclusions

There are many e-commerce sites where user's ratings are binary (like/dislike). In this paper, we examined the robustness of collaborative filtering algorithm in a binary recommender system. The results of a series of experiments demonstrate that the segment attack model has strong attacking effect against both BU-CF algorithm and BI-CF algorithm. Generally, the two algorithms have bigger hit ratios when attack size increases,

but BU-CF algorithm reflects stronger than BI-CF algorithm. As for the influence of the filler size, the segment attack affects BU-CF algorithm more when the filler size increases in the beginning, but the effect changes little after the filler size reaches a value. So it is the same with BI-CF algorithm.

Acknowledgment

This work is partially supported by National Natural Science Foundation of China (Grant Nos. 71361012 and 71363022), and by the Foundation of Jiangxi Provincial Department of Education (GJJ. 12744).

References

[1] L. Y. Lü, M. Medo, C. H. Yeung, *et al.*, "Recommender System", Physics Reports, vol. 519, no. 1, **(2012)**, pp. 1- 49.
[2] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state of the art and possible extensions", Knowledge and Data Engineering, IEEE Transactions, vol. 17, no. 6, **(2005)**, pp. 734–749.
[3] M. J. Pazzan and D. Billsus, "Content-based recommendation systems", The adaptive web, Springer, **(2007)**, pp. 325-341.
[4] M. Lipczak, Y. Hu, Y. Kollet and E. Milios, "Tag sources for recommendation in collaborative tagging systems", Proceedings ECML/PKDD Discovery Challenge, **(2009)**, pp. 157-172.
[5] I. Cantador, D. Vallet and J. M. Jose, "Measuring vertex centrality in cooccurrence graphs for online social tag recommendation", Proceedings ECML/PKDD Discovery Challenge, **(2009)**, pp. 17-33.
[6] S. Ju and K. B. Hwang, "A weighting scheme for tag recommendation in social bookmarking systems", Proceedings ECML/PKDD Discovery Challenge, **(2009)**, pp. 109–118.
[7] D. Goldberg, D. Nichols, B. M. Oki and D. Terry, "Using collaborative filtering to weave an information tapestry," Commun ACM, vol. 35, **(1992)**, pp. 61-70.
[8] J. B. Schafer, D. Frankowski, J. Herlocker and S. Sen, "Collaborative filtering recommender systems", In: The adaptive web, Springer, **(2007)**, pp. 291-324.
[9] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", Proceedings of the 14th Conference on Uncertainty in artificial intelligence. Morgan Kaufmann Publishers Inc., **(1998)**, pp. 43-52.
[10] B. Sarwar, G. Karypis, J. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th International World Wide Web Conference, **(2001)**.
[11] A. Albadvi and M. Shahbazi, "A hybrid recommendation technique based on product category attributes", Expert Systems with Applications, vol. 36, no. 9, **(2009)**, pp. 11480–11488.
[12] D. R. Liu, C. H. Lai and W. J. Lee, "A hybrid of sequential rules and collaborative filtering for product recommendation", Information Sciences, vol. 179, no. 20, **(2009)**, pp. 3505–3519.
[13] J. Salter and N. Antonopoulos, "Cinema screen recommender agent: combining collaborative and content-based filtering", IEEE Intelligent Systems, vol. 21, no. 1, **(2006)**, pp. 35–41.
[14] I. Gunes, C. Kaleli, A. Bilge and H. Polat, "Shilling attacks against recommender systems: A comprehensive survey", Artificial Intelligence Review, vol. 11, **(2012)**, pp. 1-33.
[15] P. Resnick, N. Iacovou, M. Sushak, *et al.*, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In Proceedings of ACM Conference on Computer Supported Cooperative Work, New York, USA: ACM Press, **(1994)**, pp. 175~186.
[16] X. Su and T. Khoshgoftaar, "A survey of collaborative filtering techniques", Advances in Artificial Intelligence, **(2009)**.
[17] S. Lam and J. Reidl, "Shilling recommender systems for fun and profit", in Proceedings of the 13th International WWW Conference, New York, **(2004)** May.
[18] B. Mobasher, R. Burke, C. Williams and R. Bhaumik. "Analysis and detection of segment-focused attacks against collaborative recommendation", Proceedings of the 2005 WebKDD Workshop, **(2006)**.
[19] R. Burke, B. Mobasher, C. Williams and R. Bhaumik, "Classification features for attack detection in collaborative recommender systems", Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, USA: ACM Press, **(2006)**, pp. 542- 547.
[20] R. Burke, B. Mobasher, R. Zabicki and R. Bhaumik, "Identifying attack models for secure recommendation". In Beyond Personalization Workshop at the International Conference on Intelligent User Interfaces, **(2005)**, pp. 347-361.
[21] R. Bhaumik, C. A. Williams, B. Mobasher and R. D. Burke, "Securing collaborative filtering against malicious attacks through anomaly detection", Proceedings of the 4th workshop on intelligent techniques for web personalization, Boston, MA, **(2006)**.

[22] R. D. Burke, B. Mobasher, C. A. Williams and R. Bhaumik, "Detecting profile injection attacks in collaborative recommender systems", Proceedings of the 8th IEEE conference on e-commerce technology, San Francisco, CA, USA, **(2006)**, pp. 23-30.
[23] R. D. Burke, M. P. O'Mahony and N. J. Hurley, "Robust collaborative recommendation", Recommender systems handbook. Springer, New York, **(2011)**, pp. 805-835.
[24] M. P. O'Mahony, N. J. Hurley and G. C. M. Silvestre, "Detecting noise in recommender system databases", Proceedings of the 11th international conference on intelligent user interfaces, Sydney, Australia, **(2006)**, pp. 109-115.
[25] Q. Long and Q. Hu, "Robust evaluation of binary collaborative recommendation under profile injection attack", In: Proceedings of the IEEE international conference on progress in informatics and computing, Shanghai, China, **(2010)**, pp. 1246-1250.
[26] R. Burke, B. Mobasher, C. Williams and R. Bhaumik, "Segment-based injection attacks against collaborative recommender systems", Proceedings of the International Conference of Data Mining, **(2005)**, pp. 577-580.
[27] B. Mobasher, R. Burke, C. Williams and R. Bhaumik, "Analysis and detection of segment-focused attacks against collaborative recommendation", Proceedings of the 2005 WebKDD Workshop, **(2006)**.
[28] J. S. Breese, D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", In Proceedings of the 14th conference on Uncertainty in Artificial Intelligence, **(1998)**, pp.43-52.
[29] J. G. Liu, C. X. Jia, T. Zhou, *et al.*, "Personal recommendation via modified collaborative filtering", Physica A, vol. 388, **(2009)**, pp. 462-468.
[30] R. Burke, B. Mobasher and R. Bhaumik, "Limited knowledge shilling attacks in collaborative filtering systems", Proceedings of the 3rd IJCAI Workshop in intelligent Techniques for Personalization, **(2005)**.

Authors

Fu-Guo Zhang, He was born in 1969, is an associate professor in the School of Information and Technology at Jiangxi University of Finance and Economics, Nanchang, China, where he received his Ph.D. degree in 2009. His research interests include recommender system and social network trust.

