DETECTION OF PROFILE INJECTION ATTACKS IN RECOMMENDER SYSTEMS USING SSA-BASED CHANGE-POINT ANALYSIS

Parthasarathi Chakraborty[1], Sunil Karforma[2]

Abstract: ***Profile-injection attacks are very common to recommender systems where a number of fake user profiles are inserted into the system to influence the recommendations made to the users. The researchers have identified different attack models. All of these models change the rating distribution of the target items. In this paper we have analyzed the time series of ratings of target items to identify whether any structural change occurred in the time series due to attack. We have used singular spectrum analysis based change-point detection technique to identify push/nuke attacks.***

Keywords: Singular spectrum Analysis, change-point detection, time series, singular value decomposition, profile-injection attack

1. INTRODUCTION

Many web sites attempt to help users by incorporating a recommender system that provides users with a list of items and/or web pages that are likely to interest them. Content-based filtering and collaborative filtering are usually applied to predict these recommendations. Among these two, Collaborative filtering is the most common approach for designing e-commerce recommender systems. It works by building a database of items with users' opinions on them. Then a specific user is matched against this database in order to find her neighbors, those with whom he or she shares similar tastes. As the system is open to user input, chance of attack on it is always there. The researchers have discussed different types of attacks. The ultimate target of all type of profile injection attacks are either to push a product (or a group of products) or to nuke a product (or a group of products). In case of Random Attack [1] a pre-specified rating is assigned to the target item and random ratings are assigned to the filler items whereas in average attacks [1], rating of each filler item corresponds to the mean rating for that item. Some additional attack types have been specified by Burke et. al. [2] namely Bandwagon Attack, Segment attack, Reverse Bandwagon Attack and Love/Hate Attack. The last one is a very simple attack and requires no system knowledge where the attack profile consist of minimum/ maximum rating value for target items and maximum/ minimum rating value for filler items for nuke/push attack.

2. RELATED WORKS

Detection of profile-injection attacks on recommender systems have been studied by many researchers. Supervised classification techniques have been used in [2] in order to distinguish attack profiles from genuine user profiles. Authors of paper [3] have shown that statistical process control(SPO) based approach can be effective in detecting items that are likely to be under attack.

In order to detect random attacks, several empirical metrics have been proposed and evaluated for analyzing rating patterns of attack profiles by Chirita et al. [4]. Zhang et al. [5] detects random attacks by computing the log-likelihood of each rating profile given the low-dimensional linear model that best describes the original rating matrix.

A time series based approach have been developed in paper [6], which can reveal the presence of a wide range of profile injection attacks. They have grouped a certain number of consecutive ratings into windows and compute the sample average and sample entropy in each window. They have analyzed the time series of the computed sample average and sample entropy to detect attack events. A heuristic algorithm that adaptively changes the window size has also been proposed by them form practical scenario where the number of attacks is unknown.

3. OUR APPROACH

What is common to all types of profile-injection attacks is attack to the system changes the rating distribution of the target items (and filler items in some types of attack). Like the authors of paper [6], we also consider the ratings of a particular item as a time series and analyze it for detection of attack.

In order to get benefit from attack event, the attacker must inject the false ratings of items into the system within a short span of time. In other words, more and more the attack will be distributed over time the lesser will be the degree of benefit from the attack event. If we plot the timestamp values of the ratings of the target item(s) along a timescale we must observe high concentration in the attack zone.

The main highlight of this paper is detecting structural changes in time series of ratings of the target item due to attack event. In order to find that, we try to detect change-point(s) in the time series of ratings of the target item after occurring the attack event. We applied the change-point detection algorithm developed by Moskvina et. al. [7], [15] which is based on singular spectrum analysis. In the next section we first give a brief introduction to the change-point detection strategies developed by researchers and then singular spectrum analysis of time series data. Then we produce the algorithm for change-point detection developed by Moskvina et. al. [7],[15] which has been used in the current research.

4. CHANGE-POINT DETECTION

The change-point detection is the problem of discovering time points in a time series where abrupt change in data occurs. Change-point detection is useful in many real-world applications like climate change analysis, intrusion detection in computer networks, segmentation of signals in stream data, fault analysis and many more. The problem of change-point detection has been studied for long time in the domain of statistics and several strategies have been devised by the researchers. In one approach the probability distributions of time-series samples over past and present intervals are compared and change-point is declared when the two

distributions are becoming significantly different [8]. CUSUM (cumulative sum) [8] and GLR (generalized likelihood ratio) [9,10] follow this approach where the logarithm of the likelihood ratio between two consecutive intervals in time-series data are monitored for detecting change-point. In another approach named subspace identification, the subspaces spanned by the columns of an extended observability matrix generated by a state-space model with system noise are compared for change-point detection. Kawahara et al [11] , Ide and Tsuda [12] follow this strategy. Unlike the above strategies, some non-parametric approaches are also devised for detecting change-points. In paper [13] non-parametric density estimation is used for calculating the likelihood ratio. Using Kullback-Leibler importance estimation procedure (KLIEP) [14] and singular-spectrum analysis[7], [15] in change-point detection are more recent approaches in this direction. In this paper we have used the change-point detection algorithm based on singular spectrum analysis developed by [7], [15] for detecting profile injection attacks.

4.1 Singular Spectrum Analysis : The Singular Spectrum Analysis (SSA) technique decomposes the original time series into the sum of a small number of independent and interpretable components such as a slowly varying trend, harmonic terms and a structure-less noise. It has been introduced by Broomhead and King in their paper [16].

The SSA technique has four major steps. In the first step the trajectory matrix is constructed from the time series and then in the second stage, the trajectory matrix is represented as a sum of rank-one bi-orthogonal elementary matrices using singular value decomposition(SVD) technique. The third step deals with the grouping of the selected principal components, which mostly describe the signal. In the last step, the signal is reconstructed from the group of selected components. In their paper [17] Goljandina et al. describes a generalized version of SSA.

4.2 Change-point detection algorithm based on SSA : As change-point detection is typically a sequential problem and SSA performs the analysis of the time series structure in a non-sequential manner, Moskviva et al. [7], [15] applied Singular Value decomposition(SVD) to the lag-covariance matrices computed in a sequence of time intervals, either [n+1,n+N] or [1,n+N] where n is the iteration number and N is the length of the time interval where the trajectory matrix is computed.

The outline of the change-point detection algorithm as developed by Moskviva et al [19] is-

- Select a windowed portion of the signal and apply Singular Spectrum Analysis (SSA) to it. The structure of that portion of the signal will be taken as an l-dimensional subspace by SSA.

  IF the structure changes further along the signal THEN

- The distance of trajectory matrix vectors to the subspace will increase and indicates change.

OTHERWISE

The vectors of the trajectory matrix further along the signal will stay close to this subspace and indicates no change.

The detailed steps are-

1. Consider the rating time series for the target item is $x_1, x_2, x_3, \ldots, x_N$ where $N<\infty$.

2. Choose window width m and the lag parameter M such that M<=m/2. Set K=m-M+1.

3. For each n=0,1,…,N-m-M take an interval of time series [n+1, n+m] and define the trajectory matrix Xn, size M by K given by equation (1).

$$X^{(n)} = \begin{pmatrix} x_{n+1} & x_{n+2} & \cdots & x_{n+K} \\ x_{n+2} & x_{n+3} & \cdots & x_{n+K+1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n+M} & x_{n+M+1} & \cdots & x_{n+m} \end{pmatrix} \quad (1)$$

4. For each n define the lag-covariance matrix,

$$R_n = \frac{1}{K}\left(X^{(n)}\right)\left(X^{(n)}\right)^T \quad (2)$$

5. Apply SVD on R_n to get M eigenvalues and eigenvectors. Sort the eigenvalues in decreasing order.

6. Select the number of components, L to use for change-point detection in such a way that the selected group of components represent most of the signal.

7. Two parameters of test interval p and q (both greater than 0) will be chosen.

8. Compute the detection $D_{n,L,p,q}$ statistics, as

$$D_{n,L,p,q} = \sum_{j=p+1}^{q} \left(X_j^{(n)} \left(X_j^{(n)}\right)^T - \left(X_j^{(n)}\right)^T PP^T X_j^{(n)} \right) \quad (3)$$

where $P_1, P_2, \ldots\ldots, P_L$ are the L selected eigenvectors and P is the $M \times L$ matrix with columns $P_1, P_2, \ldots\ldots, P_L$. The part of the sample $x_{n+1}, x_{n+2}, x_{n+3}, \ldots, x_{n+M}$

is used to construct the trajectory matrix $X^{(n)}$ and termed as 'training sample' and the rest part $x_{n+p+1}, x_{n+p+2}, x_{n+p+3}, \ldots, x_{n+q+M-1}$ is used to construct the vectors $X_j^{(n)}$ for $j = p+1, \ldots, q$ for computing the sum of squared distances, $D_{n,L,p,q}$ is termed as 'validation sample' [19]. The local minima of the $D_{n,L,p,q}$ function preceding its large values is computed for identifying the change-point locations.

9. an additional CUSUM statistic is computed for $n = 0, \ldots, N-m-M$

$$W_1 = S_1, \tag{4}$$

$$W_{n+1} = \max\left[0, W_n + S_{n+1} - S_n - k/\sqrt{M(q-p)}\right], \; n >= 1 \text{[15]}$$

where k is a small nonnegative constant, a reasonable value is $k = 1/\left(3\sqrt{M(q-p)}\right)$ and $S_n = D_{n,L,p,q}/v_n$, v_n is an estimator [19] of the normalized sum of squared distances, $D_{n,L,p,q}$ at time intervals, at which the hypothesis of no change can be accepted.

If W_n exceeds a threshold as defined in [19] then the change-point estimate is a first point with non-zero value of Wn before reaching this threshold [19].

5. RESULT

In our experiment we have used MovieLens dataset (movielens.umn.edu). The data set used contained 100,000 ratings from 943 users and 1682 movies (items), with each user rated at least 20 items. The item sparsity is easily computed as 0.9369. The ratings in the MovieLens dataset are integers ranging from 1 to 5. The software used in change-point detection is developed by Moskvina et. al.(available at http://www.cf.ac.uk/maths /subsites/stats/changepoint/). Based on no of ratings and average ratings, items have been categorized into five groups.

1. Items with largest no of ratings (Category-I).
2. Items with highest average ratings (can be called as popular items). Here only those items were considered which have been rated by more than 30 users (Category-II).
3. Items with highest average ratings as well as have large no of ratings (greater than 300 ratings) (Category-III).
4. Items with lowest average ratings (less than 2.5) but have large no of ratings (greater than 100 ratings) (Category-IV).

5. Items with lowest average ratings (less than 2.5) and small no of ratings(minimum 50 ratings) (Category-V).

Table 1 shows item no and their related statistics for the first two items from each category.

Item no	Category	Average rating	No of ratings
49	I,III	4.35	583
257	I	3.8	509
962	II	4.29	41
250	II	4.26	46
99	III	4.15	508
121	IV	2.34	106
242	IV	2.44	132
930	V	2.15	57
449	V	2.39	63

Table 1. Average rating and no of ratings of first two items from each category.

The aim of this categorization of items is to see how well the change-point detection algorithm [15] detects attack event. In case of push attack, the maximum rating value (i.e. 5 in MovieLens dataset) has been injected into the system and in case of nuke attack the minimum rating value (i.e. 1 in MovieLens dataset) has been injected. The attack size has been considered as a percentage of the no of ratings in the time series of the target item. Figure 1 shows the time series for target item 49 after push attack of size 5%. The corresponding detection statistics and CUSUM statistics has been shown in the middle and lower part of figure 1. The first three principal components have been taken here. The red bar in the figure shows the detection of attack event.

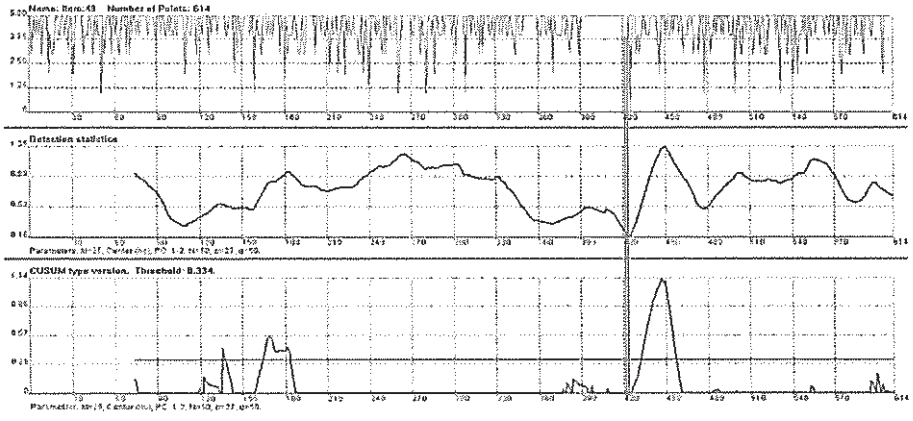

Fig. 1. Time series for item 49 after 5% push attack and the corresponding Detection statistics and CUSUM statistics.

When there is no change in the structure of the time series both detection statistics show low values and a large increase in their values indicate attack event. The attack event shown in the figure is computed from the CUSUM statistic. The first time location where the statistic is equal to zero preceding the peak value is searched for and defined as attack event.

It is to be noted that item 49 has largest no of ratings and at the same time has highest average rating value. So, in case of push attack it is difficult to detect the attack event. Bellow 5% attack size, the detection of attack event is not clear using the algorithm [15] we are using. In the contrary, nuke attack of size 1% of the no of ratings is easily identified by the algorithm [15] as shown in figure 2.

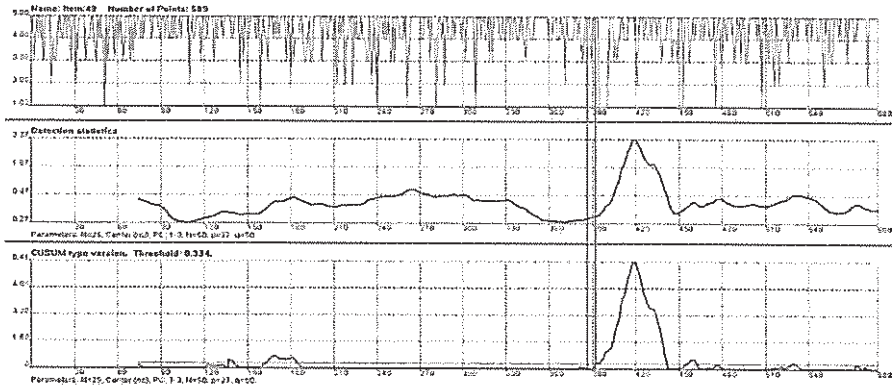

Fig. 2. Time series for item 49 after 1% nuke attack and the corresponding Detection statistics and CUSUM statistics.

For the items with very low average rating and a sufficiently large no of ratings, detection of push attack is comparatively easier than detection of nuke attack. Figure 3 shows the detection of push attack of size 2% for item 121 having average rating as low as 2.34 with 106 no of ratings. Figure 4 shows the detection of nuke attack of size 5% for the same item, but bellow 5% attack size, attack event can not be detected clearly.

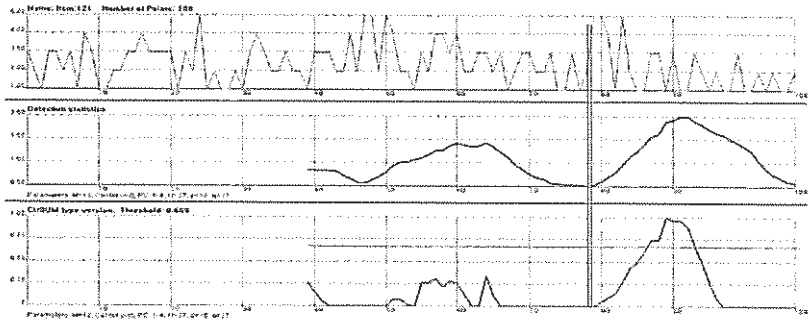

Fig. 3. Time series for item 121 after 2% push attack and the corresponding Detection statistics and CUSUM statistics.

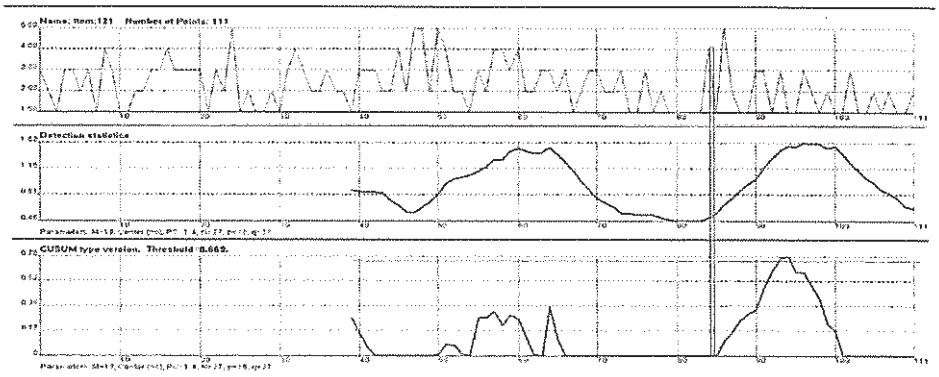

Fig. 4. Time series for item 121 after 5% nuke attack and the corresponding Detection statistics and CUSUM statistics.

Finally we have considered item no 930 as a candidate item having very low average rating (2.15) and very small no of ratings (57) also. In case of nuke attack, detection of attack event is not clear bellow attack size 10%. For push attack also, 5% attack size is not sufficient for getting clear indication about attack event. This is due to the fact that for item no 930, 5% means 2.85. i.e. less than three ratings. So two consecutive maximum rating value or minimum rating value may be present in the time series in normal course also.

6. REFERENCES

1. Lam, S. And Riedl, J. Shilling recommender systems for fun and profit. In Proceedings of the 13th International WWW Conference (New York, NY)(2004).
2. Burke, R.,Mobasher, B.,Williams, C., And Bhaumik, R. 2006b. Detecting profile injection attacks in collaborative recommender systems. In Proceedings of the IEEE Joint Conference on Ecommerce Technology and Enterprise Computing, E-Commerce and E-Services (CEC/EEE 2006, Palo Alto, CA)(2006).
3. R. Bhaumik, C. Williams, B. Mobasher, and R. Burke. Securing collaborative filtering against malicious attacks through anomaly detection. In Proceedings of the 4th Workshop on Intelligent Techniques for Web Personalization (ITWP'06), at AAAI'06, Boston, (2006).
4. P.-A. Chirita, W. Nejdl, and C. Zam_r. Preventing shilling attacks in online recommender systems. In Proc. of ACM Int. Workshop on Web Information and Data Management, pages 67-74,(2005).
5. S. Zhang, J. Ford, and F. Makedon. Analysis of a low-dimensional linear model under recommendation attacks. The 29th Annual International ACM Conference on Research & Development on Information Retrieval (SIGIR 2006), Seattle, WA, August 6-11,(2006).
6. Zhang S, Chakrabarti A, Ford J, Makedon F. Attack detection in time series for recommender systems. In: KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 809–814(2006).
7. Moskvina, V., Zhigljavsky, A. An algorithm based on singular spectrum analysis for change-point detection. Communication in Statistics: Simulation and Computation, vol. 32, No. 2, pp. 319-352,(2003).
8. Basseville, M. and Nikiforov, V., Detection of Abrupt Changes: Theory and Application, Prentice-Hall, Inc., Englewood Cliffs, N. J., (1993).
9. Gustafsson, F., The Marginalized Likelihood Ratio Test for Detecting Abrupt Changes, IEEE Trans. On Automatic Control, 41(1): 66-78, (1996).

10. Gustafsson, F., Adaptive Filtering and Change Detection, John Wiley & Sons Inc., (2000).
11. Y. Kawahara, T. Yairi, and K. Machida. Change-point detection in time-series data based on subspace identification. In Proceedings of the 7th IEEE International Conference on Data Mining, pages 559–564,(2007).
12. T. Ide and K. Tsuda. Change-point detection using Krylov subspace learning. In Proceedings of the SIAM International Conference on Data Mining, pages 515–520,(2007).
13. Brodsky, B. and Darkhovsky, B., Nonparametric Methods in Change-Point Problems, Kluwer Academic Publishers,(1993).
14. M. Sugiyama, T. Suzuki, S. Nakajima, H.i Kashima, P. von Buenau, and M. Kawanabe. Direct importance estimation for covariate shift adaptation. Annals of the Institute of Statistical Mathematics, 60(4):699–746,(2008).
15. V. Moskvina and A. Zhigljavsky. Change-point detection algorithm based on the singular spectrum analysis. Communications in Statistics: Simulation and Computation, 32: 319–352,(2003).
16. Broomhead, D. S. and King, G. P. Extracting qualitative dynamics from experimental data. Physica D 20, 217-236(1986).
17. Goljandina N., Nekrutkin V., Zhigljavsky A Analysis of Time Series Structure: SSA and related techniques, London; Chapman and Hall,(2001).
18. Takens, F. Detecting strange attractors in turbulence, pp. 366 in Dynamical Systems and Turbulence (D. A. Rand and L. S. Young, editors). Berlin: Springer,(1981).
19. Moskvina, V. Application of the singular spectrum analysis for change-point detection in time series. PhD Thesis, Cardiff University, UK, (2001).

****

---

[1]*Author 1: University Institute of Technology, The University of Burdwan*
psc755@gmail.com
[2]*Author 2: Department of Computer Science, The University of Burdwan*
sunilkarforma@yahoo.com