ISSN: 2321-7782 (Online)
Volume 2, Issue 12, December 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Article / Survey Paper / Case Study
Available online at: www.ijarcsms.com

A Survey on Profile-injection attacks in Recommender Systems using Outlier Analysis

Jiten Harishbhai Dhimmar[1]
Research Scholar
Computer Engineering Dept
Parul Institute of Technology
Vadodara – India

Raksha Chauhan[1]
Assistant Professor
Computer Science & Engineering Dept
Parul Institute of Technology
Vadodara – India

Abstract: ***E-Commerce recommender systems are vulnerable to different types of profile-injection attacks where a number of fake user profiles are inserted into the system to influence the recommendations made to the users. The ultimate target of all type of profile-injection attacks is either to push or nuke a product. We are focusing on comparing different clustering based outlier detection techniques likes PAM, CLARA, CLARANS and ECLARANS for detecting such attack profiles. In all this algorithms, the attack profiles are considered as outlier in user rating datasets. ECLARANS algorithm improves accuracy and reduces the time complexity of detecting outliers among different clustering techniques for outlier detection. ECLARANS algorithm gives best result for identification of the attack profiles in recommender System.***

Keywords: Recommender System, Profile-injection attack, Outlier Detection, PAM, CLARA, CLARANS, ECLARANS.

I. INTRODUCTION

Many web sites incorporate a recommender system in order to help users by providing a list of items that are likely to interest them from a huge database of items. In other words, recommender systems help to overcome the problem of information overload on the Internet by providing personalized recommendations to the users [1].

Content-based and collaborative filtering are two main approaches used in designing recommender systems. In content-based recommender systems, items are recommended based on the content of the items and target user's ratings. Collaborative Filtering generates recommendations for a given user by considering opinions of other users. A specific user is matched against other users in the rating database in order to find her neighbors-users with similar tastes. Collaborative Filtering has been used successfully by many e-commerce sites and in the area of information filtering [1]. Collaborative recommender systems are known to be highly vulnerable to profile injection attacks, attacks that involve the insertion of biased profiles into the ratings database for the purpose of altering the system's recommendation behaviour [2].

II. ATTACKS ON RECOMMENDER SYSTEMS

People are frequently required to make choices about items without having a significant knowledge of the range of choices available. Consequently, we often seek recommendations from others relating to which movies to see, which books to read or which car to buy etc. Collaborative recommendation algorithms operate in a similar fashion and can be used to filter information and recommend personalized content that satisfy the particular needs and tastes of individual users[3]. These algorithms have been successfully employed in many online settings and work by gathering preference data and opinions from users and using this information to generate recommendations for others.

While people are relatively adept at assessing the reliability of friends and associates and valuing recommendations from such sources accordingly, it is much more difficult to make judgments concerning users of online environments given their anonymous or pseudo-anonymous nature. Since it is practically impossible to determine in advance the motivations and integrity

of those who use online systems [3], there is no guarantee that the preferences expressed for items reflect the true opinions of users.

In addition, it is often feasible to create a number of identities within a single system and thus the potential for shilling attacks or profile injection attacks to occur exists. These attacks involve the creation of multiple attack profiles which are typically designed to reflect the true preferences of genuine users for certain items, while the target item is assigned a biased rating with the intention of promoting or demoting recommendations made for the item in question. Such attacks are referred to as product push and product nuke attacks, respectively [3]. Since it has been demonstrated that the presence of even small quantities of attack profiles can significantly bias recommendations, it is vital that online systems are protected against these kinds of attack.

In Random Attack a pre-specified rating is assigned to the target item and random ratings are assigned to the filler items whereas in average attacks, rating of each filler item corresponds to the mean rating for that item. Some additional attack types have been specified by Burke 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 consists of minimum/ maximum rating value for target items and maximum/ minimum rating value for filler items for nuke/push attack [4].

III. Outlier Detection

In literature, the researchers have proposed several outlier detection techniques. They can be broadly categorized into different groups namely distance based approach, density based approach, clustering based approach and depth based approach. In clustering based approach, the clusters having small number of members are considered as the clusters consisting of outliers assuming that outliers are a small percentage of the total data. The main advantage of this approach over the other approaches is that the outlier detection is totally unsupervised [4]. Moreover, clustering-based techniques are capable of being used in an incremental mode [5].

There are various kinds of Clustering based outlier detection approach which are following:

A. PAM (Partition around Medoid)

PAM uses a k-medoid method for clustering. It is very robust when compared with k-means in the presence of noise and outliers. Mainly in contains two phases Build phase and Swap phase.

Build phase: This step is sequentially select k objects which is centrally located. This k objects to be used as k-medoids.

Swap phase: Calculates the total cost for each pair of selected and non-selected object [6].

B. CLARA (Clustering Large Applications)

CLARA is introduced to overcome the problem of PAM. This works in larger data set than PAM. This method takes only a sample of data from the data set instead of taking full data set. It randomly selects the data and chooses the medoid using PAM algorithm [6].

C. CLARANS (Clustering Large Applications Based on Randomized Search)

This method is similar to PAM and CLARA. It starts with the selection of medoids randomly. It draws the neighbour dynamically. It checks “maxneighbour” for swapping. If the pair is negative then it chooses another medoid set. Otherwise it chooses current selection of medoids as local optimum and restarts with the new selection of medoids randomly. It stops the process until returns the best [6].

D. ECLARANS (ENHANCED CLARANS)

This method is different from PAM, CLARA and CLARANS. Thus method is produced to improve the accuracy of outliers. ECLARANS is a new partitioning algorithm which is an improvement of CLARANS to form clusters with selecting proper arbitrary nodes instead of selecting as random searching operations. The algorithm is similar to CLARANS but these selected arbitrary nodes reduce the number of iterations of CLARANS [6].

A comparative study of the above algorithm has been done and concludes that ECLARANS is best technique among them. The clustering technique ECLARANS has detects more number of outlier compare with PAM, CLARA, and CLARANS [6]. Thus it could be shows that the ECLARANS algorithm improves accuracy of detecting the outliers. CLARANS takes lesser amount of time to detect the outlier means that CLARANS reduce time complexity when compared with other clustering algorithms for outlier detection [5] & [6].

IV. CONCLUSION

E-Commerce recommender systems are vulnerable to different types of profile-injection attacks where a number of fake user profiles are inserted into the system to influence the recommendations made to the users. ECLARANS algorithm improves accuracy of detecting outliers from different clustering based outlier detection techniques and time complexity of ECLARANS algorithm is less then PAM and CLARA algorithms. So, an ECLARANS algorithm is the best algorithm compared with PAM, CLARA and CLARANS algorithms for the detection of attack profiles in E-Commerce recommender system.

ACKNOWLEDGEMENT

I thanks to Assistant Professor Raksha Chauhan who is from Computer & Science Department of Parul Institute And Technology, Vadodara. She is also my guide who inspire me for doing such work and due to their support here I am able to write this survey paper and she also provide me good knowledge about outlier detection techniques and Guide me for doing this work.

References

1. Partha Sarathi Chakraborty, "A Scalable Collaborative Filtering Based Recommender System Using Incremental Clustering", IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009.
2. Burke R., Mobasher B., Williams C. and Bhaumik R., "Detecting profile injection attacks in collaborative recommender systems", In Proceedings of the 8th IEEE International Conference on E-Commerce Technology and the 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services (CEC/EEE'06), 2006.
3. M. O., K. Bryan and P. Cunningham, "Unsupervised retrieval of attack profiles in collaborative recommender systems," in Technical Report, University College Dublin, 2008.
4. Parthasarathi Chakraborty, Sunil Karforma, "Detection of Profile-injection Attacks in Recommender Systems Using Outlier Analysis", In:Procedia Technology, Volume 10, 2013.
5. Deepak Sinwar, Dr. Sudesh Kumar, "Study of Different Clustering Approaches for Outlier Detection", IJCSC, Volume 4, Number 2 September 2013.
6. S.Vijayarani, S.Nithya, "An Efficient Clustering Algorithm for Outlier Detection", International Journal of Computer Applications (0975 – 8887), Volume 32– No.7, October 2011.
7. Al- Zoubi, M. B., "An Effective Clustering-Based Approach for Outlier Detection", European Journal of Scientific Research, Vol. 28, No. 2, 2009, pp. 310-316.
8. H.-P. Kriegel, M. S. hubert, and A. Zimek. "Angle-based outlier detection in highdimensional data". In KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages444–452, New York, NY, USA, 2008. ACM.
9. Chad A.Williams, Bamshad Mobasher, Robin Burke, "Defending recommender systems: detection of profile injection attacks", SOCA (2007).
10. Ghazaleh Aghili, Mehdi Shajari, Shahram Khadivi, Mohammad Amin Morid, "Using Genre Interest of Users to Detect Profile Injection Attacks in Movie Recommender Systems ". In Proceeding of IEEE International Conference on Machine Learning and Applications, 2011.
11. Loureiro,A., L. Torgo and C. Soares, "Outlier Detection using Clustering Methods: a Data Cleaning Application", in Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector. Bonn, Germany, 2004.
12. Jiawei Han, Micheline Kamber & Jian pei, Data Mining - Concept and techniques (3rd ad). Elsevier 2012.

AUTHOR(S) PROFILE

Jiten Harishbhai Dhimmar, received the B.E. degree in Information Technology from Government Engineering College, Sec-28, Ghandhinagar is affiliated with Gujarat Technological University (GTU). Currently he is doing his M.E. Degree from Parul Institute of Technology, Vadodara is affiliated with Gujarat Technological University(GTU) and his interesting area of research is in Outlier Detection which is refer as a sub-area of Data Mining.