International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Probabilistic Inference and Trustworthiness Evaluation of Associative Links toward Malicious Attack Detection for Online Recommendations

*1Ishrath Nousheen, 2A Anishka, 2K Mamatha

1Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

*2,3Student, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

Article Info
Volume 9, Issue 5
Page Number : 630-633

Publication Issue
September-October-2022
Pages : 606-611

Article History
Accepted : 10 Oct 2022
Published : 30 Oct 2022

ABSTRACT

Today, spamming mails is one of the biggest issues faced by everyone in the world of the Internet. In such a world, email is mostly shared by everyone to share the information and files because of their easy way of communication and for their low cost. But such emails are mostly affecting the professionals as well as individuals by the way of sending spam emails. Every day, the rate of spam emails and spam messages is increasing. Such spam emails are mostly sent by people to earn income or for any advertisement for their benefit. This increasing amount of spam mail causes traffic congestion and waste of time for those who are receiving that spam mail. The real cost of spam emails is very much higher than one can imagine. Sometimes, the spam emails also have some links which have malware. And also, some people will get irritated once they see their inbox which is having more spam mails. Sometimes, the users easily get trapped into financial fraud actions, by seeing the spam mails such as job alert mails and commercial mails and offer emails. It may also cause the person to have some mental stress. To reduce all these risks, the system has proposed a machine learning model which will detect spam mail and non-spam emails, and also this system will optimize the data by removing the unwanted mails which contain the advertisement mails and also some useless emails and also some fraud mails. This proposed system will detect the spam mails and ham emails with the dataset consisting of spam mails and after identifying spam mails this system will remove that spam emails and this proposed system will calculate the amount of storage before and after the removal of spam mails.

Keywords : Machine Learning, Attack Detection, Recommendation.

I. INTRODUCTION

Recommender systems (RS)—which are extensively deployed in various systems such as e-commerce, social networks, search engines, news portals, hiring platforms, intelligent assistants, smart home and smart city services, as well as healthcare and financial applications—have been acknowledged for their capacity to deliver high-quality services that bridge the gap between users and items by delivering tailored content for each individual. recommender systems not only help users to find relevant information more efficiently, but also directly influence the human decision-making process by providing relevant suggestions or even shape users' worldviews by exposing users to the selected content. Overall, recommender system is the frontier of Human-centered AI research and works as the bridge between humans and AI. However, for every plus there is a minus, RS may offer both promise and perils. There are growing concerns that the irresponsible use of recommendation techniques may bring counter-effects and untrustworthy issues, such as compromised user trust due to non-transparency, unfair treatment of different users, producers, or platforms, privacy concerns due to extensive use of user's private data for personalization, echo chambers due to the lack of controllability for users that leads to repeated reinforcement of users' existing interests — the list just continues to expand. These vulnerabilities significantly limit the development and deployment of recommendation algorithms and may even lead to severe economic and social issues. As a result, only considering recommendation accuracy is not enough when developing modern recommendation systems. We also need to make sure that the models are fair, have not been tampered with, will not fall apart in different conditions, and can be understood by humans. Moreover, the design and development process of RS also needs to be transparent and inclusive. All of these considerations beyond accuracy that makes recommender systems safe, responsible, and worthy of our trust are related to trustworthy recommender systems research. Since recommender system is an important direction of Human-centered AI research that directly involves humans in the loop, Trustworthy Recommender System (TRS) has been leading the research of Trustworthy Artificial Intelligence (TAI) over the past years on various perspectives, such as the definition, method and evaluation of trustworthiness on explainability, fairness, robustness, and privacy, as well as how humans interact with trustworthy AI systems. Therefore, as a vital instantiation of trustworthy AI under the context of recommender system research, in this survey, we introduce trustworthy recommender systems as competent RS that incorporates the core aspects of trustworthiness such as explainability, fairness, privacy, robustness and controllability. We believe that incorporating these aspects when designing recommender systems will improve their responsibility, gain trust from human users, and significantly promote recommender systems for social good.

II. RELATED WORK

Group Shilling Attacks. To escape from the existing methods of detecting individual shilling attacks (e.g., random attack, average attack, and AOP attack), Wang et al. proposed two generative models of group shilling attack, called GSAGens and GSAGenl. In these attack models, fake profiles are first generated based on one type of individual shilling attacks. Based on which, group shilling attack profiles are then constructed and injected into a set of genuine profiles. The GSAGens model has more stringent conditions when generating group shilling profiles; hence, the group size under GSAGens is smaller than that under GSAGenl. Considering the attack effect on the target items, we only use GSAGenl to generate the group shilling attack profiles, in which the fake profile includes the selected item set, the filler item set, the target item set, and the unrated item set. More details for the group shilling attacks used in this paper are described as follows:

(1) GSAGenl Ran: generate loose group attack profiles based on a random attack, where the selected items are null, the filler items are randomly chosen, and only one attacker from the whole group rates the items. The filler item rating is the system mean. The target item rating is set to rmax or rmin

(2) GSAGenl Ave: generate loose group attack profiles on the basis of an average attack, where the selected items are null, the filler items are randomly chosen, and only one attacker from the whole group rates the items. The filler item rating is the item mean. The target item rating is set to rmax or rmin

(3) GSAGenl AOP: generate loose group attack profiles based on 50% AOP attack, where the selected items are null, the filler items are randomly chosen, and only one attacker from the whole group rates those items with top 50% popularity. The filler item rating is the item mean. The target item rating is set to rmax or rmin

(4) GSAGenl GOAT: generate loose group attack profiles based on the adversarial attack, called GOAT [11], where each fake user's selected and filler items are randomly chosen from an item-item graph based on genuine user profiles. A generative adversarial network is used to generate the ratings of the selected and filler items based on the genuine rating distribution. The target items have ratings of rmax – 1 or rmin + 1

(5) GSAGenl Mixed: generate mixed multiple shilling groups generated according to the four abovementioned group shilling attacks.

III. PROPOSED SYSTEM

Depicts the two stages of the KC-GCN detection framework: influential user extraction and attack user identification. In the first stage, we build the user relationship graph by determining the user similarity based on the item suspicious time window. Next, we use the k-clique community discovery algorithm to generate suspicious candidate groups. Finally, we obtain the influential users by calculating the user nearest-neighbor similarity. In the second stage, we extract the users' initial embeddings from four dimensions.We then combined the extracted user initial embeddings with the structural features hidden in the user relationship graph to train a semi-supervised classifier based on a twolayer GCN, which only utilizes the labels of the identified influential users.

Constructing a Weighted User Relationship Graph. Attackers in a shilling group typically cooperate to quickly enhance or demote the recommendation of one or more target items. Based on this characteristic of group attacks, the rating distribution of a target item may fluctuate during the attacked time period. Therefore, we construct a weighted user relationship graph by extracting the suspicious time windows of the suspicious items and calculating the correlation between users within the suspicious time windows.

IV. CONCLUSION AND FUTURE WORK

In this work, we put forward a two-stage semi-supervised model to validly detect various types of group shilling attacks on recommender systems. First, we construct a user relationship graph and spot the influential users. In the graph, the edge weight is calculated by analyzing the user similarity over suspicious time intervals on each item. Next, we generate the initial user embeddings based on the proposed four indicators describing the behavior difference between attack and genuine users. A GCN-based classifier is trained, and the attack users are detected based on the influential user labels. The experimental results prove the effectiveness and the generality of KC-GCN.

In the future work, we will automatically determine the labels of most influential users by further analyzing the structural properties of the weighted user relationship graph. We will also study the multiaspect data to further help identify users of group shilling attack.

V. REFERENCES

[1]. H. Li, K. Wang, Y. Sun, and X. Mou, “Application of recommendation systems based on deep learning,” in Recent Challenges in Intelligent Information and Database Systems. ACIIDS 2021, Communications in Computer and Information Science, T. P. Hong, K. Wojtkiewicz, R. Chawuthai, and P. Sitek, Eds., pp. 85–97, Springer, Singapore, 2021.

[2]. Q. Shambour, “A deep learning based algorithm for multicriteria recommender systems,” Knowledge-Based Systems, vol. 211, article 106545, 2021.

[3]. N. Nassar, A. Jafar, and Y. Rahhal, “A novel deep multi-criteria collaborative filtering model for recommendation system,” Knowledge-Based Systems, vol. 187, no. 7, pp. 104811.1– 104811.7, 2020.

[4]. Y. Feng, F. Lv, W. Shen et al., “Deep session interest network for click-through rate prediction,” 2019, https://arxiv.org/abs/ 1905.06482.

[5]. F. Lv, T. Jin, C. Yu et al., “SDM: sequential deep matching model for online large-scale recommender system,” in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2635–2643, Beijing, China, 2019.

[6]. S. Zhang, H. Liu, J. He, S. Han, and X. Du, “Deep sequential model for anchor recommendation on live streaming platforms,” Big Data Mining and Analytics, vol. 4, no. 3, pp. 173–182, 2021.

[7]. P. Nitu, J. Coelho, and P. Madiraju, “Improvising personalized travel recommendation system with recency effects,” Big Data Mining and Analytics, vol. 4, no. 3, pp. 139–154, 2021.

[8]. T. Li, C. Li, J. Luo, and L. Song, “Wireless recommendations for internet of vehicles: recent advances, challenges, and opportunities,” Intelligent and Converged Networks, vol. 1, no. 1, pp. 1–17, 2020.

[9]. H. Li, M. Gao, F. Zhou, Y. Wang, Q. Fan, and L. Yang, “Fusing hypergraph spectral features for shilling attack detection,” Journal of Information Security and Applications, vol. 63, article 103051, 2021.

Cite this article as :

Ishrath Nousheen, A Anishka, K Mamatha, "Probabilistic Inference and Trustworthiness Evaluation of Associative Links toward Malicious Attack Detection for Online Recommendations", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 9 Issue 5, pp. 630-633, September-October 2022.

Journal URL : https://ijsrst.com/IJSRST2182944