Detecting Anomalous Ratings in Collaborative Filtering Recommender Systems

Zhihai Yang, Xi’an Jiaotong University, Xi’an, China
Zhongmin Cai, Xi’an Jiaotong University, Xi’an, China

ABSTRACT

Online rating data is ubiquitous on existing popular E-commerce websites such as Amazon, Yelp etc., which influences deeply the following customer choices about products used by E-businessman. Collaborative filtering recommender systems (CFRSs) play crucial role in rating systems. Since CFRSs are highly vulnerable to “shilling” attacks, it is common occurrence that attackers contaminate the rating systems with malicious rates to achieve their attack intentions. Despite detection methods based on such attacks have received much attention, the problem of detection accuracy remains largely unsolved. Moreover, few can scale up to handle large networks. This paper proposes a fast and effective detection method which combines two stages to find out abnormal users. Firstly, the manuscript employs a graph mining method to spot automatically suspicious nodes in a constructed graph with millions of nodes. And then, this manuscript continue to determine abnormal users by exploiting suspected target items based on the result of first stage. Experiments evaluate the effectiveness of the method.

KEYWORDS

Attack Detection, Collaborative Filtering Recommender Systems, Graph Mining, Intrusion Detection, Shilling Attack

1. INTRODUCTION

Personalization recommender systems (RSs) become more and more popular in some well-known E-commerce websites such as Amazon, eBay etc. (Bryan, OMahony, & Cunningham, 2008; Burke, Mobasher, & Williams, 2006; Mehta, Hofmann, & Fankhauser, 2007; Zhang, & Kulkarni, 2014). These E-commerce services have not gained higher customer satisfaction on products or services but also derived more benefits from customer ratings since being successfully armed with personalized recommendations. Collaborative filtering recommender systems (CFRSs) have been proved to be one of the most of RSs. However, CFRSs are highly vulnerable to “shilling” attacks or “profile injection” attacks since its openness, it is common occurrence that attackers contaminate the CFRSs with malicious rates to achieve their attack intentions. To avoid negative results, the companies either defend such attacks or benefit from customer ratings by detecting functional weakness as well as deficiencies of products for improving the detected lacks (Gnnemann, Gnnemann, & Faloutsos, 2014a; Gnnemann, Gnnemann, & Faloutsos, 2014b). Thus, constructing an effective detection method to defend such attacks and remove them from the CFRSs is crucial.

Despite detection methods based on such attacks have received much attention, the problem of detection accuracy remains largely unsolved. For example, it is difficult to extract more effective
features to characterize the difference between attack profiles and genuine profiles (Burke, Mobasher, & Williams, 2006; Zhang, & Kulkarni, 2014; Williams, Mobasher, & Burke, 2007; Williams, Mobasher, Burke, & Bhaumik, 2007; Morid, & Shajari, 2014; Mehta, 2007). Furthermore, some features based on calculating similarity between users are effective to capture the concerned attackers, but the computation time is unreasonable, especially for large-scale networks within million of nodes (Bryan, OMahony, & Cunningham, 2008; Burke, Mobasher, & Williams, 2006; Mehta, Hofmann, & Fankhauser, 2007; Zhang, & Kulkarni, 2014; Williams, Mobasher, & Burke, 2007; Williams, Mobasher, Burke, & Bhaumik, 2007; Morid, & Shajari, 2014; Mehta, 2007). It is noteworthy that few focus on the real world datasets including Amazon, Yelp etc, to discover or detect the hidden attacks (Gnnemann, Gnnemann, & Faloutsos, 2014a; Gnnemann, Gnnemann, & Faloutsos, 2014b).

Given a graph within million of nodes (consists of attackers and genuine users), how can we detect automatically anomalous or suspicious nodes with reasonable computation time? In practice, the attackers are paid to make certain accounts seem more legitimate or famous through giving them many additional users. The attackers deliver these purchases through either generating fake accounts or controlling real accounts through malware and using them to follow their “customers” (Jiang, Cui, Beutel, Faloutsos, & Yang, 2014). In this case, the attackers are manipulating well-designed graph to give certain accounts undue credibility, since the attackers only require adding edges to the graph.

In this paper, we propose a fast and effective detection method which consists of two stages to find out abnormal users as far as possible. Firstly, we employ a graph mining method to spot automatically suspicious nodes in a given graph with million of nodes (Jiang, Cui, Beutel, Faloutsos, & Yang, 2014). The goal of this stage is expect to quickly find out suspected users as many as possible. It is noteworthy that both the ratings and items rated by users are equal important to determine the edges between nodes for constructing a graph. Since we just use the items rated by users to determine the edges in the fist stage, it is difficult to fully capture the concerned attackers. Based on this situation, we continue to capture the potential attackers by exploiting suspected target items based on remained users from the first stage. As is known, the attackers will target one or more specific items with lowest or highest rating many times if they want to demote (called nuke attack) or promote (called push attack) the items to the recommendation lists, we can decide the suspected target items by using an absolute count threshold (Zhou, Koh, Wen, Burki, & Dobbie, 2014). We evaluate the efficiency of our method on MovieLens-100K dataset with diverse attack models.

The rest of this paper is organized as follows. Section 2 reviews some related work. Section 3 shows attack models and attack profiles. In Section 4, we detail our proposed method. In Section 5, experimental results are reported and analyzed. Finally, we conclude the paper with a brief summary and prospect the directions of future works.

2. RELATED WORK

The main objective of our work is to spot abnormal users in CFRSs for defending “shilling” attacks or “profile injection” attacks. In the following, we discuss methods related to our work. Su, Zeng, & Chen, 2005 presented a spreading similarity algorithm in order to capture groups of similar attackers. Then, Burke, Mobasher, & Williams, 2006; Zhang, & Kulkarni, 2014; Williams, Mobasher, & Burke, 2007; Williams, Mobasher, Burke, & Bhaumik, 2007; Morid, & Shajari, 2014; Mehta, 2007 developed different attributes derived from user profiles for their utility in attack detection. One of their attributes is Degree of Similarity with Top Neighbors (DegSin) which calculates the similarity between users by using Pearson Correlation Coefficient. Mehta, Hofmann, & Fankhauser, 2007 proposed an unsupervised detection method based on principal component analysis and performed
A Novel Cue based Picture Word Shape Character Password Creation Scheme
www.igi-global.com/article/a-novel-cue-based-picture-word-shape-character-password-creation-scheme/134053?camid=4v1a