Expert Systems with Applications 41 (2014) 3671–3681

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Robustness analysis of privacy-preserving model-based recommendation schemes

Alper Bilge, Ihsan Gunes, Huseyin Polat *

Computer Engineering Department, Anadolu University, 26470 Eskisehir, Turkey

ARTICLE INFO

Keywords:
Robustness
Shilling
Privacy
Recommendation
Model
Collaborative filtering

ABSTRACT

Privacy-preserving model-based recommendation methods are preferable over privacy-preserving memory-based schemes due to their online efficiency. Model-based prediction algorithms without privacy concerns have been investigated with respect to shilling attacks. Similarly, various privacy-preserving model-based recommendation techniques have been proposed to handle privacy issues. However, privacy-preserving model-based collaborative filtering schemes might be subjected to shilling or profile injection attacks. Therefore, their robustness against such attacks should be scrutinized.

In this paper, we investigate robustness of four well-known privacy-preserving model-based recommendation methods against six shilling attacks. We first apply masked data-based profile injection attacks to privacy-preserving *k*-means-, discrete wavelet transform-, singular value decomposition-, and item-based prediction algorithms. We then perform comprehensive experiments using real data to evaluate their robustness against profile injection attacks. Next, we compare non-private model-based methods with their privacy-preserving correspondences in terms of robustness. Moreover, well-known privacy-preserving memory- and model-based prediction methods are compared with respect to robustness against shilling attacks. Our empirical analysis show that couple of model-based schemes with privacy are very robust.

1. Introduction

With the emergence of the Internet and following developments in the Internet technologies, e-commerce has become very popular. Online shopping sites offer numerous products. Thus, customers using these e-vendors have too many product options to choose and purchase. There are various tools for customers to guide them through e-commerce web sites. Recommender systems are one of such customer friendly tools. One technique that is widely used by recommender systems is called collaborative filtering (CF), which enables user preference estimation in the recommendation systems (Birtolo & Ronca, 2013; Wen, Fang, & Guan, 2012). Recommendation schemes provide different gains such as customer constancy, increasing advertisement revenues, higher amount of sales, and so on (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013). Hence, online vendors benefit from e-commerce web sites through recommendation techniques. Similarly, CF schemes help customers by listing products that they might like among millions of available items. Thus, users waste less time for searching products in the system.

CF process consists of three major steps. They are similarity calculation, neighborhood formation, and recommendation estimation. Since the main objective of CF is to procure connection between individuals and current online data, calculating similarities, selecting neighbors, and estimating predictions become practical. In a typical CF system, an $n \times m$ user-item matrix is formed. Within this matrix, n and m symbolize the number of users and products, respectively. The matrix includes users' preferences about various products. An active user (a), who aims to get a prediction for a target item q, sends her ratings to the system. A prediction (p_{aq}) is estimated by choosing the most resembling users (Mazurowski, 2013).

Along with their advantages, CF schemes have their own challenges. Some of the main problems are listed as scalability, accuracy, privacy, and shilling (Bobadilla et al., 2013; García-Cumbreras, Montejo-Ráez, & Díaz-Galiano, 2013; Schafer, Frankowski, Herlocker, & Sen, 2007). When the system holds a huge number of users and/or items, traditional CF algorithms will suffer from serious scalability problems. For solving the scalability problem, model-based recommendation algorithms are usually preferred such as applying dimensionality reduction techniques, singular value decomposition (SVD), clustering, or item-based algorithms. Due to sparse nature of collected data, CF systems might not be able to provide sufficiently accurate predictions,

* Corresponding author. Tel.: +90 222 321 3550x6554.
E-mail addresses: abilge@anadolu.edu.tr (A. Bilge), ihsang@anadolu.edu.tr (I. Gunes), polath@anadolu.edu.tr (H. Polat).

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.11.039

which is referred to as the accuracy problem. Compared to model-based schemes, memory-based ones provide more accurate referrals. Hence, they are preferred for overcoming the accuracy problem. Customers usually do not want to disclose their preferences about various products and rated items to CF schemes. However, many recommender systems fail to protect such data. This problem is called the privacy problem. Hence, not being able to achieve data confidentiality is another disadvantage of CF systems (Calandrino, Kilzer, Narayanan, Felten, & Shmatikov, 2011). A CF system without enough privacy protection hinders data procurement or generates wrong information. Customers look for a system, which is secure in terms of individual data protection. Less qualified user data causes poorer recommendation and less accurate estimations. In a system with well-developed privacy measures, collecting qualified and dependable data is easier. Protecting individual data is possible through some techniques named privacy-preserving collaborative filtering (PPCF) (Polat & Du, 2005).

Malicious users or parties might insert fake profiles into the CF system's database to manipulate the estimated predictions on behalf of their advantages. This phenomenon is also known as shilling. CF systems might be defenseless against shilling attacks (Burke, Mobasher, Bhaumik, & Williams, 2005a; O'Mahony, Hurley, & Silvestre, 2006). Fake user profiles created by malevolent users, vendors, or rivals can be used to defraud system recommendation. The main purpose behind these attacks is to manipulate the outputs of the system. For instance, while some attacks aim at misleading users towards particular items, others try to decrease popularity of certain products (O'Mahony, Hurley, & Silvestre, 2005). Thus, preventing the recommendation system from shilling attacks requires specific detection methods (Williams, Mobasher, & Burke, 2007; Zhang, Luo, Weng, & Li, 2009) and robust recommendation algorithms (Burke, O'Mahony, & Hurley, 2011; Jia, Zhang, & Liu, 2013; Noh, Kang, Oh, & Kim, 2014).

PPCF schemes are grouped as memory- or model-based schemes. Memory-based techniques with privacy are the simplest heuristic methods. It is easy to use such methods for producing predictions. Since memory-based algorithms work online, it is easy to add a new user or product into the collection. It is not required to evaluate the content of the recommended items. The mechanism scales well with co-rated items. On the other hand, the size of data can be a disadvantage for scaling those systems. When a new user enters into the system, recommendation for that user might not be possible due to data sparseness. Privacy-preserving model-based CF algorithms produce a model relying on user ratings as well as providing predictions. Even though they perform better in terms of scalability and sparsity problems, their implementation is harder compared to memory-based ones. They find item or user similarities off-line via the model. When a new item or user is added, a fresh model should be created; however, this process is computationally expensive. Also, as useful data can be lost during a specific model production, accuracy may decrease.

CF schemes without privacy concerns are investigated in terms of profile injection attacks. Different approaches are proposed to handle the shilling problem. Similarly, numerous PPCF schemes are suggested to overcome the privacy problem. In addition to protecting confidentiality, preventive methods for PPCF schemes against shilling attacks are also claimed. However, there are not sufficient studies to scrutinize PPCF schemes with respect to shilling attacks. Researches on PPCF's robustness against shilling attacks are not high in number. There are couple of studies only in the literature that investigate shilling attacks against memory-based PPCF techniques (Gunes, Bilge, Kaleli, & Polat, 2013; Gunes, Bilge, & Polat, 2013). In these previous studies, two memory-based algorithms were studied to show how robust they were against these attacks. In this study, the question of whether or not model-based PPCF schemes are robust against shilling attacks is examined. Robustness of four state-of-the-art model-based PPCF schemes is examined against six attack models, which are designed to manipulate private preference collections. Investigated model-based schemes are *k*-means-, SVD-, item-, and discrete wavelet transform (DWT)-based PPCF schemes. Modified versions of random, average, bandwagon, and segment push attacks along with reverse bandwagon and love/hate nuke attack models are applied against such PPCF schemes. Relying on algorithms' recommendation mechanism and experimental outcomes, their robustness against profile injections are discussed.

The contributions of this article in general are listed below:

1. We first apply six shilling attack models to four model-based PPCF schemes.
2. Comprehensive real data-based experiments are conducted to evaluate the robustness of the four model-based PPCF algorithms against the six attack models.
3. Those four model-based PPCF schemes are compared with their non-private correspondences and a well-known memory-based PPCF scheme in terms of robustness against the six shilling attack models.

The remainder of the paper is structured as follows. In Section 2, related studies are reviewed and the differences between this work and the existing ones are briefly presented. In Section 3, preliminary works are described. We describe six attack models that are implemented against PPCF schemes in Section 4. In Section 5, real data-based experiments and their results are given. Finally, in Section 6, conclusions and future work are presented.

2. Related work

Model-based CF schemes have been proposed to enhance online efficiency of the filtering systems. Wen and Zhou (2012) proposed an improved item-based prediction algorithm on dynamic item clustering method. The authors introduced a similitude threshold model to divide item space into clusters dynamically. Kim, Kwon, Cho, and Kang (2011) developed a recommendation system for suggesting TV genres using *k*-means clustering and ontology technique. They showed the feasibility of their proposed scheme using real TV viewing history. Russell and Yoon (2008) suggested to apply DWT on recommender systems. Data are transformed on these systems and reduced significantly to enhance the amount of time for making a prediction. A new algorithm based on incremental SVD and generalized Hebbian algorithm was proposed by Polezhaeva (2011). The new algorithm effectively updates user/item profiles when a new user or a new item appears. It also does not require to store the initial data matrix. Bobadilla et al. (2013) offered an extensive survey to be used for CF techniques. They introduced CF tasks and their basic challenges as well as probable solutions. They then provided three main CF technique categories, which are called memory-based, model-based, and hybrid CF algorithms.

In today's world, protecting personal data has gained great importance in recommendation systems. Polat and Du (2005) used randomized perturbation techniques (RPTs) to overcome the problem named privacy on CF. In another study, Polat (2006) provided predictions on item- and SVD-based CF algorithms with privacy assured by using RPTs. Bilge and Polat (2012) discussed the ways to make successful predictions on reduced space by utilizing DWT without violating individual users' privacy. They also presented the way to apply *k*-means and bisecting *k*-means clustering on CF schemes while preserving the confidentiality of users (Bilge & Polat, 2013a; Bilge & Polat, 2013b). Parra-Arnau, Rebollo-Monedero, and Forné (2012) proposed an architecture that