Engineering 4 (2018) 21–28

Contents lists available at ScienceDirect

Engineering

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

Research
Cybersecurity—Review

Toward Privacy-Preserving Personalized Recommendation Services

Cong Wang [a,b,*], Yifeng Zheng [a,b], Jinghua Jiang [a,c], Kui Ren [d]

[a] Department of Computer Science, City University of Hong Kong, Hong Kong, China
[b] City University of Hong Kong, Shenzhen Research Institute, Shenzhen, Guangdong 518057, China
[c] Department of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China
[d] Institute of Cyber Security Research, Zhejiang University, Hangzhou, Zhejiang 310058, China

ARTICLE INFO

Article history:
Received 21 June 2017
Revised 26 September 2017
Accepted 12 February 2018
Available online 16 February 2018

Keywords:
Privacy protection
Personalized recommendation services
Targeted delivery
Collaborative filtering
Machine learning

ABSTRACT

Recommendation systems are crucially important for the delivery of personalized services to users. With personalized recommendation services, users can enjoy a variety of targeted recommendations such as movies, books, ads, restaurants, and more. In addition, personalized recommendation services have become extremely effective revenue drivers for online business. Despite the great benefits, deploying personalized recommendation services typically requires the collection of users' personal data for processing and analytics, which undesirably makes users susceptible to serious privacy violation issues. Therefore, it is of paramount importance to develop practical privacy-preserving techniques to maintain the intelligence of personalized recommendation services while respecting user privacy. In this paper, we provide a comprehensive survey of the literature related to personalized recommendation services with privacy protection. We present the general architecture of personalized recommendation systems, the privacy issues therein, and existing works that focus on privacy-preserving personalized recommendation services. We classify the existing works according to their underlying techniques for personalized recommendation and privacy protection, and thoroughly discuss and compare their merits and demerits, especially in terms of privacy and recommendation accuracy. We also identity some future research directions.

1. Introduction

Nowadays, recommendation systems are increasingly gaining popularity and are widely deployed for online services. The widespread use of recommendation systems allows users to enjoy diverse personalized recommendations for movies, books, ads, restaurants, hotels, and more. Meanwhile, personalized recommendation services have also become extremely effective revenue drivers for online business. For example, recent research suggests that 35% of what consumers purchase on Amazon and 75% of what they watch on Netflix are attributable to personalized recommendations [1]. A study conducted by the research firm Marketing Sherpa showed that 11.5% of the revenue generated in the shopping sessions of involved e-commerce sites comes from purchases of products via personalized recommendation [2].

To support personalized recommendation, the common practice of existing systems usually involves either collaborative filtering-based (CFB) recommendation or content-based (CB) recommendation [3]. CFB recommendation systems usually recommend items based on the similarity between users. For example, a user rating for a movie would be predicted based on the ratings/decisions of other similar users (classified via some metric). CB recommendation systems typically generate recommendations by comparing the properties of items with those of users' personal preference/behavioral data. For example, an ad network may compare the keywords associated with ads with the keywords indicating a user's preference in order to serve personalized ads. To obtain personalized recommendations from these systems, users are typically required to provide their personal data to the recommender for processing and analytics.

Although personalized recommendation is greatly beneficial, directly exposing users' private data to the recommender poses privacy risks for users [4–6]: ① The provided data undesirably discloses the users' personal interests to the recommender; ② the provided data may be abused by the recommender, for example, by a recommender selling user data to third parties for financial incentives without user consent [4]; and ③ the provided data

* Corresponding author.
E-mail address: congwang@cityu.edu.hk (C. Wang).

https://doi.org/10.1016/j.eng.2018.02.005

may be stolen by motivated attackers due to security breaches on the recommender side [5,6]. Therefore, it is of paramount importance to develop privacy-preserving techniques for recommendation systems, so that the intelligence of recommendation systems is preserved while user privacy is respected.

In this paper, we survey the literature related to personalized recommendation services with privacy protection. We first present the general architecture of real-world recommendation systems, and the privacy issues therein. We then provide a comprehensive survey of existing solutions that can support privacy-preserving personalized recommendation services. As mentioned above, the mechanisms adopted in recommendation systems are usually either CFB or CB. Based on this observation, we first classify the existing solutions into two broad categories: ① privacy-preserving CFB recommendation and ② privacy-preserving CB recommendation. In the first category, existing works are further classified into private neighborhood-based approaches and private machine learning-based approaches, according to the concrete plaintext techniques adopted. In the second category, existing works are further classified into private targeted advertising and private targeted coupon delivery, according to the concrete application settings. Therefore, there are four explicit categories in total.

When describing the representative existing works in each category, our key insight is to further classify them based on their underlying security strategies/techniques for privacy protection; for example, some works rely on cryptographic techniques such as homomorphic encryption and garbled circuits (GCs), while others resort to data obfuscation techniques. Among the large volume of works in this trending area, we carefully select highly cited representative works that describe popular techniques, as well as papers that deliver significantly new and emerging techniques. Our goal is to cover each category as comprehensively as possible in order to call for further motivated research activities.

The rest of this paper is organized as follows. Section 2 presents the general architecture and privacy issues of recommendation systems. Section 3 describes existing works on privacy-preserving CFB recommendation. Section 4 describes existing works on privacy-preserving CB recommendation. Section 5 discusses some future research directions, and Section 6 concludes the paper.

2. Recommendation systems

2.1. System model

Recommendation systems aim to provide accurate recommendations for users by collecting and processing their personal data using effective approaches [7]. The system model of a personalized recommendation service is illustrated in Fig. 1. It contains two primary entities: users and recommender. Each user has some personal data on her local device (e.g., a smartphone), which indicates her personal interests/preferences. The recommender collects the users' personal data, processes the collected data, and provides personalized recommendations for the users. The generated recommendations can be provided or shown to users in various ways, such as through messages and pop-up windows.

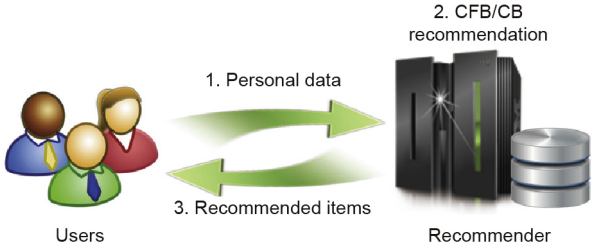

Fig. 1. The system model of a personalized recommendation service. The recommender may adopt either collaborative filtering-based (CFB) techniques or content-based (CB) techniques.

To process the collected user data for recommendation, the recommender may adopt different kinds of techniques. Roughly speaking, according to the adopted techniques, recommendation systems can be classified into two categories: CFB recommendation systems and CB recommendation systems. As mentioned earlier, CFB recommendation systems recommend items based on the similarity between users. That is, items recommended to a particular user are those preferred by other users that share similar preferences [3]. In contrast, CB recommendation systems conduct recommendation based on the properties of items, which may be described by certain explicit features (e.g., attributes and characteristics).

To leverage the similarity between users, CFB recommendation systems usually adopt either neighborhood-based approaches or machine learning-based approaches. Neighborhood-based approaches directly compute the similarity relationship between users [8], and leverage this relationship to generate personalized recommendations. In contrast, machine learning-based approaches first learn a mathematical model from the collected user data, and then apply the model to generate personalized recommendations.

2.2. Privacy issues

The more personal data a recommender collects, the more accurate recommendations users can obtain. The user data collected by the recommender may include information about the users' identity, demographic profile, behavioral data, purchase history, rating history, and more [9]. Such information can be very privacy-sensitive. For example, the demographic profile refers to demographic characteristics of the customer, such as age, gender, weight, and level of education; behavioral data refers to dynamic data of the customer, such as location, activity status, and browsing history; and rating history refers to the votes that the customer has provided on items. Providing such information to the recommender in the clear would pose undesirable privacy risks. For example, user data could be sold to a third party by the recommender without user consent, or could even be stolen by motivated attackers. Therefore, protecting user data in recommendation systems is of critical importance.

3. Privacy-preserving CFB recommendation

CFB recommendation systems typically recommend items based on similarity measures between users [3]. To support the functionality of CFB recommendation while preserving user privacy, a number of works on privacy-preserving CFB recommendation have been undertaken. Because CFB recommendation systems usually adopt either neighborhood-based approaches or machine learning-based approaches, we classify these existing works into two categories: private neighborhood-based approaches and private machine learning-based approaches.

3.1. Private neighborhood-based approaches

Existing solutions under the umbrella of private neighborhood-based approaches usually adopt techniques from two main categories. The first category is cryptographic techniques [4,6,10,11] and the second is randomization techniques [12–14]. Cryptographic technique-based solutions generally require high