Guiding Us Throughout a Sea of Data - A Survey on Recommender Systems and Its Privacy Challenges

Xiwei Wang, Marcelo Sztainberg
Computer Science Department
X-Wang9@neiu.edu, M-Sztainberg@neiu.edu

Over the past decades, the Internet has served as the backbone connecting people to others, places and things. With the sheer volume of information generated everyday, people can feel overwhelmed when having to make a selection among the multiple options that normally come up after a search or application request. For example, when searching for news articles regarding a particular topic, the search engine will present a number of results to you. When looking for some product on shopping websites, there are usually several pages of results that match the keywords. It can be very challenging for people to find their most expected information in the era of big data.

A recommender system is a program that utilizes algorithms to learn users' preferences from historical data, and predict their future interests. Recommender systems are employed everywhere in the cyberspace. Many websites including Amazon, eBay, YouTube, Facebook, Netflix, and others, have integrated automatic personalized recommendation techniques into their systems, in order to help users find their most desired information.

While recommender systems have become a common feature on most web applications and sites, one of the major issues around its use is privacy concerns. A regular recommender system requires the users to share their online behavior data, such as their past shopping records, browsing history, visited places, so that it can learn their preferences. This can potentially deter people from using the system because these data are considered as users' privacy and many do not feel comfortable sharing the information with other parties.

In this research, we studied several recommendation algorithms, and compared their performance as well as prediction accuracy on real-world datasets. We also proposed a novel nonnegative matrix factorization (NMF) based privacy-preserving point-of-interest recommendation framework, in which the latent factors in NMF are learned on user group preference instead of individual user preference. Recommendations are made by personalizing the group preference on user's local devices. There are no location or check-in data collected from the users anywhere throughout the learning and recommendation processes. Some preliminary results on a regular recommender system were established and two GUI applications were developed. The on-going research focuses on integrating the privacy-preserving framework into the system and verifying the effectiveness as well as the recommendation accuracy of the proposed model.