Application of Differential Privacy to Recommendation Systems

ABSTRACT

The deluge of data on the web, online products, and services has made recommendation systems an integral part of the Web realm. They are employed in a wide range of applications starting from eCommerce sites, through the Social Web, to healthcare apps. Recommenders are leveraged to enhance the product sales, to help user in quick decision making, and to suggest relevant products to users from massive product catalogue. Typically, recommender systems rely on users' personal information to train the system, to prioritize the relevant items to a specific user based on the previous preferences and to predict the rating for new products based on the user's behavior. However, such data usage in recommender systems hampers the user's sensitive information and poses a severe threat to individual privacy.

This thesis addresses the problems of privacy preservation in recommender systems using differential privacy. Differential privacy features a provable privacy guarantee, but challenges the application of the same in the recommender systems. The following are some of the challenges in applying differential privacy to recommender:

- The dataset's sparsity aids attackers to launch various attacks on the recommenders
- High dimensionality of data in recommenders in turn results in large noise addition when perturbation based techniques are used
- Maintaining a balance between privacy and utility in perturbation based techniques.

To address these challenges, the present research work aims at devising and analysing privacy-preserving recommendation techniques using differential privacy. This thesis analyses and evaluates five different

differentially private algorithms on recommendation system. The predictive accuracy of the proposed algorithms is compared with the existing private and non-private techniques. Further, the utility and privacy of proposed algorithms are theoretically analysed. The following algorithms are proposed to overcome the challenges mentioned:

- **Safe ℓ injection algorithm:** This algorithm imputes the missing ratings and also addresses sparsity with differentially private imputation improve the recommendation results. It comes with a dual advantage of privacy protection and effective recommendation. Since the missing ratings are imputed, the user preferences are protected from the adversary. In addition, an optimal imputation of uninteresting items reduces the sparsity which inherently results in better recommendation.

- **Random ALS Perturbation:** In matrix factorization, the user and the item factors are used to predict the results. Existing private algorithms perturbs all the factors by noise addition which results in poor performance. The proposed random Alternating Least Square (ALS) perturbation algorithm chooses a random factor for perturbation. As a result, the algorithm minimizes the amount of noise that is added for privacy and increases the accuracy.

- **DeepPriv:** This algorithm is designed to preserve privacy in deep neural networks. When the naive differential privacy technique is applied on high dimensional embeddings, accuracy is reduced. This issue is overcome by DeepPriv. An adaptive perturbation is designed for optimal noise addition to the embeddings of the deep neural network. Such adaptive perturbation particularly for high dimensional data is not available in the existing works.

- **Private Bloom:** An efficient local differentially private algorithm (LDP) has to address issues including transmission cost, convergence time and amount of noise addition. Since LDP assures the users raw data remains in the user’s device, it needs a large amount of noise when compared to all the other privacy techniques. These issues are addressed by proposing a private bloom filter using Laplace mechanism.

CryptoDP: Finally, a hybrid algorithm is proposed to protect the rating matrix as well as the collaborative filtering model. Hybrid algorithm called CryptoDP utilizes the advantages of cryptography and differential privacy. Such a comprehensive protection mechanism needs noise addition at multiple points of the system. However, the suggested solution does noise addition only once during the similarity matrix formation process and uses cryptography to address privacy at other points in the system.