Amrita Vishwa Vidyapeetham
Amrita School of Engineering, Coimbatore
Department of Computer Science and Engineering
15CSE401 - Machine Learning & Data Mining
Case Study

Title: PrivRank

Team Name: Orion **Group Number: 5**

S.No	Name	Roll Number
1	Abhishek S	CB.EN.U4CSE17003
2	Maheshaa Ganapath GL	CB.EN.U4CSE17038
3	Manojkumar V K	CB.EN.U4CSE17040
4	Sudharshan S	CB.EN.U4CSE17059
5	Vasudevan P	CB.EN.U4CSE17067

Abstract:

A personalized recommendation is crucial to help users find pertinent information. It often relies on a large collection of user data, in particular users' online activity (e.g., tagging/rating/checking-in) on social media, to mine user preference. However, releasing such user activity data makes users vulnerable to inference attacks, as private data (e.g., gender) can often be inferred from the users' activity data. PrivRank is a customizable and continuous privacy-preserving social media data publishing framework protecting users against inference attacks while enabling personalized recommendations. Its key idea is to continuously obfuscate user activity data such that the privacy leakage of user-specified private data is minimized under a given data distortion budget, which bounds the ranking loss incurred from the data obfuscation process to preserve the utility of the data for enabling recommendations. An empirical evaluation on both synthetic and real-world datasets shows that our framework can efficiently provide effective and continuous protection of user-specified private data, while still preserving the utility of the obfuscated data for the personalized ranking-based recommendation.

Introduction:

Developing effective recommendation engines is critical in the era of Big Data to provide pertinent information to the users. To deliver high-quality and personalized recommendations, online services such as e-commerce applications typically rely on a large collection of user data, particularly user activity data on social media, such as tagging/rating records, comments, check-ins, or other types of user activity data. In practice, many users are willing to release the data (or data streams) about their online activities on social media to a service provider in exchange for getting high-quality personalized recommendations. Such user activity data are referred to as public data. However, they often consider part of the data from their social media profile as private, such as gender, income level, political view, or social contacts. Those data are referred to as private data. Although users may refuse to release private data, the inherent correlation between public and private data often causes serious privacy leakage.

Existing Work:

To protect user privacy when publishing user data, the current practice mainly relies on policies or user agreements, e.g., on the use and storage of the published data. However, this approach cannot guarantee that the users' sensitive information is protected from a malicious attacker. Therefore, to provide effective privacy protection when releasing user data, privacy-preserving data publishing has been widely studied. Its key idea is to obfuscate user data such that published data remain useful for some application scenarios while the individual's privacy is preserved. According to the attacks considered, existing work can be classified into two categories. The first category is based on heuristic techniques to protect ad-hoc defined user privacy. The second

category is theory-based and focuses on the fact that published data should provide attackers with as little additional information as possible beyond background knowledge.

Challenges:

To release private data, the inherent correlation between public and private data often causes serious privacy leakage. It is thus crucial to protect user private data when releasing public data to recommendation engines.
More distortion of public data leads to better privacy protection, as it makes it harder for adversaries to infer private data. Also reconstruction of values poses serious challenges.

Literature Survey

Dingqi Yang et. al. [1] introduced a customizable and continuous privacy-preserving social media data publishing framework. Bounding the ranking loss is incurred from the data obfuscation process using the Kendall-τ rank distance and cluster-wise obfuscation function. This framework can be extended by considering the data types with continuous values rather than discretized values and exploring further data utility beyond personalized recommendation.

Dong Li et. al. [2] propose an improved collaborative filtering algorithm and community detection algorithm. They select a part of user communities from the user network projected by the user-item network as the candidate neighboring user set for the target user, thereby reducing calculation time and increasing recommendation speed and accuracy of the recommendation system. The newly proposed method has a perfect combination of social network technology and collaborative filtering technology, which greatly increases the recommendation system performance.

A. Zhang et. al. [3] proposes a privacy-preserving, personalized and relevant content recommendation video consumption system. This model uses convex optimization to learn a probability mapping from actual ratings to perturbed ratings that minimize distortion subject to a privacy constraint. This could also be adapted to protect privacy in the context of social networks: users could be informed of the privacy risks of actions such as likes, connecting to friends. In such a context, data distortion could amount to simply avoiding to take some actions, or avoiding the release of some data.

D.Yang et. al. [4] in their paper, proposed a spatial-temporal activity preference (STAP) model to enable a wide range of ubiquitous applications, such as personalized context-aware location recommendation and group-oriented advertisement. They evaluated their proposed approach on three real-world datasets collected from New York and Tokyo, and showed that STAP consistently outperforms the baseline approaches in various settings.

L. Wang et. al. [5] put forth a location privacy-preserving task allocation framework with geo obfuscation to protect users' locations during task assignments. A mixed-integer nonlinear programming problem formulated to minimize the expected travel distance of the selected workers under the constraint of differential privacy and outperforms Laplace obfuscation.

S. Salamatian et. al. [6] evaluated the correlations between political views and TV viewing habits. In this research, the data distortion is done according to a privacy-preserving probabilistic mapping. They map by solving a convex optimization problem, which minimizes information leakage under a distortion constraint. They had a practical challenge: the optimization can become intractable and face scalability issues when data assumes values in large size alphabets or is high dimensional. They first reduce the optimization size by a quantization step and show how to generate privacy mappings under quantization. This demonstrates that good privacy properties can be achieved with limited distortion so as not to undermine the original purpose of the publicly released data, e.g. recommendations.

L. Sweeney et. al. [7] provides a formal protection model named k-anonymity and a set of accompanying policies for deployment. A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. This paper also examines re-identification attacks that can be realized on releases that adhere to k-anonymity unless accompanying policies are respected. The k-anonymity protection model is important because it forms the basis on which the real-world systems known as Datafly, μ-Argus and k-Similar provide guarantees of privacy protection.

Benjamin C. M. Fung et. al. [8] studied different approaches that have been developed and conducted a survey. In this work, the data is being anonymized so that the attacker is denied from earning useful information. They introduce the term QID which consists of a set of attributes that could potentially identify a user. This also helps in knowing the record linkage details. They use this to anonymize the information by reducing the number of linkages which in term prevents attackers from mining patterns. They found out that the data publishers find it difficult in choosing QID.

Rakesh Agrawal et. al. [9] were one of the earlier pioneers of data mining research and were, in fact, responsible for identifying that the direction of it would be fruitful only if privacy concerns are incorporated in the techniques that are developed. In this work, they propose that they let users provide modified values of sensitive attributes. In this method, the values of an attribute are partitioned into a set of disjoint mutually-exclusive classes. Then the value is distorted by considering two distributions, namely the uniform and the gaussian distribution. The data is then reconstructed using the inverse functions of the distributions.

S. Banerjee et. al. [10] discusses the possibility of a recommendation system to be accurate if the end-user does not trust the system and proposes a new algorithm where this can be made possible. They propose a new algorithm which takes a new approach for the information-scarce regime (where each user rates only a vanishing fraction of the total item set).

A. Evfimievski et. al. [11] discusses the disadvantages of randomization of aggregate data to preserve privacy. They present a new formulation of privacy breaches, together with a methodology, "amplification", for limiting them. Unlike earlier approaches, amplification makes it possible to guarantee limits on privacy breaches without any knowledge of the distribution of the original data.

C. Dwork et. al. [12] studies the addition of random noise to user data to preserve privacy and tries to prove that privacy can be preserved by calibrating the standard deviation of the noise according to the sensitivity of the noise function. Roughly speaking, this is the amount that any single argument to noise function can change its output. The new analysis shows that for several applications, substantially lesser noise was required than what was previously understood to be the case.

Proposed Work:

Differential Privacy is essentially a property that the system should maintain, rather than a specific way of calculating. Therefore, the framework designed includes different perturbation methods to carry out predictions in a differentially private manner. The simple technique of noise addition is fully fit for predicting the scores while protecting original ratings without leakage. It offers a mathematical definition of privacy and a provable privacy guarantee for each record in the dataset. Intuitively, the output of the computation should not reveal too much information about any record in the dataset. The probability of the output is insensitive to small input changes, whether one record is in the dataset or not. Differential Privacy is presented in a series of papers, mainly used in data publishing and data mining.

Definition: A randomized computation K satisfies ε-DP if, for any neighboring datasets A and B differing on at most one record, and for all subsets of possible outputs $S \subseteq Range(K)$.

$$P[K(A) \in S] \leq exp(\varepsilon) \times P[K(B) \in S]$$

where ε is the privacy budget to make the trade-off between privacy and accuracy. The value of ε is generally set to a small positive value. The smaller it is, the higher privacy and lower accuracy it provides and vice versa.

Dataset Description:

GroupLens Research [13] has collected and made available rating data sets from the MovieLens web site (http://movielens.org). The data sets were collected over various periods, depending on the size of the set. These files contain 1,000,209 ratings of approximately 3,900 movies made by 6,040 MovieLens users who joined MovieLens in 2000.All demographic information is provided voluntarily by the users and is not checked for accuracy. Only users who have provided some demographic information are included in this data set.

The dataset is available to the general public and can be downloaded at: MovieLens Dataset

Implementation:

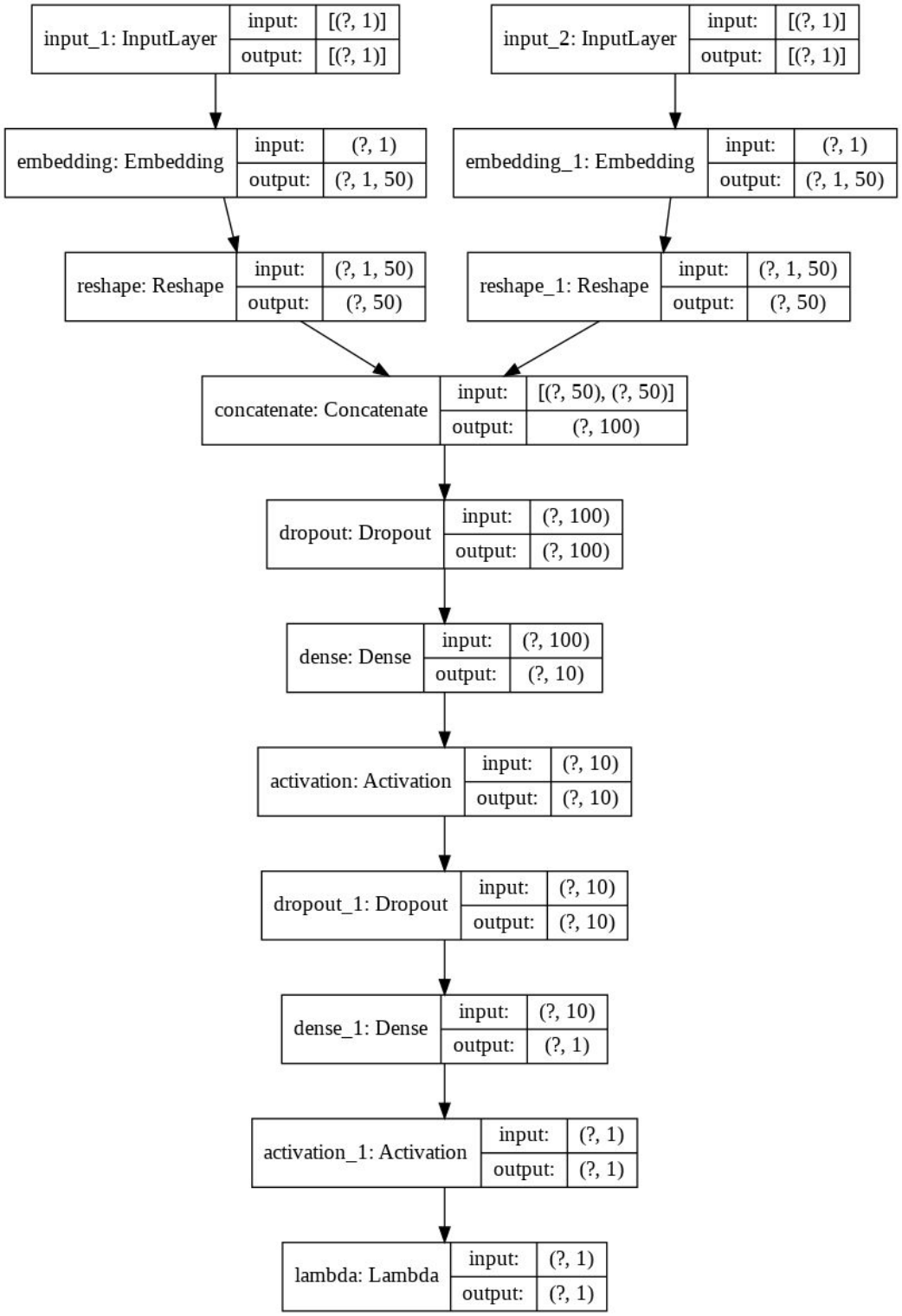
In this case study, we have used neural network based collaborative filtering. The objective is to be able to predict ratings for movies a user has not yet watched. The movies with the highest predicted ratings can then be recommended to the user. The steps in the model are as follows:

- Map user ID to a "user vector" via an embedding matrix
- Map movie ID to a "movie vector" via an embedding matrix
- Compute the dot product between the user vector and movie vector, to obtain the a match score between the user and the movie (predicted rating).
- Train the embeddings via gradient descent using all known user-movie pairs.

Inorder to maintain user-privacy, two different approaches were followed:

- By directly adding noise to the training data so that the exact user details and preferences are masked. To experiment with noise scales, four different noise scales were selected, which are 0.5, 1.0, 1.5 and 2.0. These noises were added to the ratings. Four identical models were trained for each of the noisy ratings and results were obtained.
- By applying differential privacy using the tensorflow-privacy tool in Python while training. The basic idea of the used approach, called differentially private stochastic gradient descent (DP-SGD), is to modify the gradients used in stochastic gradient descent (SGD), which lies at the core of almost all deep learning algorithms. Models trained with DP-SGD provide provable differential privacy guarantees for their input data. TensorFlow Privacy provides code that wraps an existing TensorFlow optimizer to create a variant that implements DP-SGD.

The following is the architecture of the neural network that has been used to train the model:



Results:

The following are the results obtained

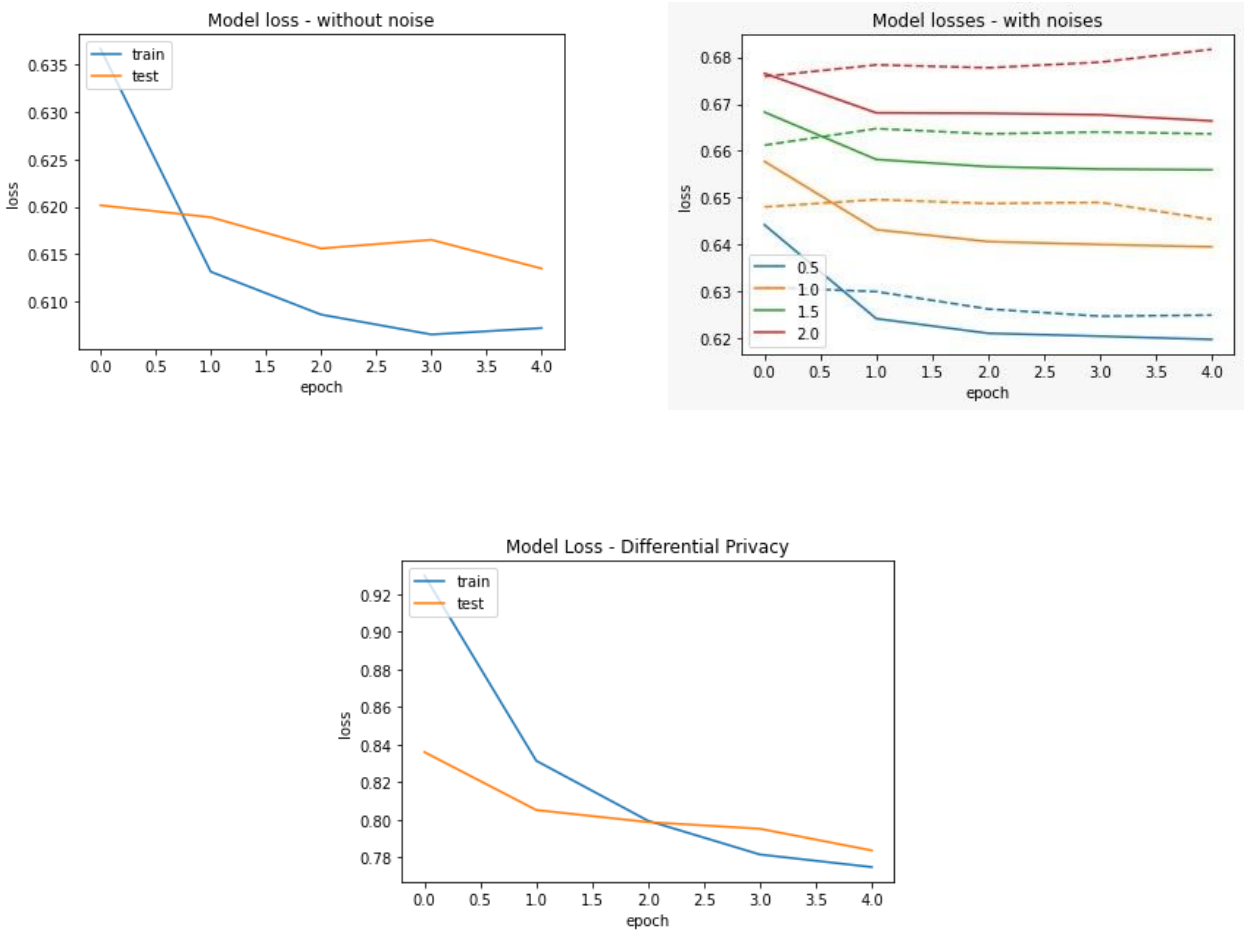


Figure: *Comparison of results between different approaches*

The used approach satisfies differential privacy with eps = 0.623 and delta = 1e-05.

Conclusion and Future Work:

From the above results, it is evident that the loss is highest in the network trained using differential privacy and the lowest when noiseless dataset is used, implying that as the level of security increases, there is an inherent increase in the loss. However, for the Movielens dataset, even though there is a small difference in the validation loss, the recommendations given to the user are almost similar in the context that the recommended movies are of similar genres.

There are other privacy preserving algorithms like K-Anonymity which should be researched upon. The use of federated learning on top of such algorithms is bound to have better recommendations with reduced loss.

References:

[1] D. Yang, B. Qu and P. Cudré-Mauroux, "Privacy-Preserving Social Media Data Publishing for Personalized Ranking-Based Recommendation," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 3, pp. 507-520, 1 March 2019, DOI: 10.1109/TKDE.2018.2840974.

[2] Xiaofeng Li, Dong Li, "An Improved Collaborative Filtering Recommendation Algorithm and Recommendation Strategy", Mobile Information Systems, vol. 2019, Article ID 3560968, 11 pages, 2019. https://doi.org/10.1155/2019/3560968

[3] Zhang, A., S. Bhamidipati, N. Fawaz and Branislav Kveton. “PriView: Media Consumption and Recommendation Meet Privacy Against Inference Attacks.” (2014).

[4] D. Yang, D. Zhang, V. W. Zheng and Z. Yu, "Modeling User Activity Preference by Leveraging User Spatial-Temporal Characteristics in LBSNs," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 45, no. 1, pp. 129-142, Jan. 2015, DOI: 10.1109/TSMC.2014.2327053.-

[5] Leye Wang, Dingqi Yang, Xiao Han, Tianben Wang, Daqing Zhang, and Xiaojuan Ma. 2017. Location Privacy-Preserving Task Allocation for Mobile Crowdsensing with Differential Geo-Obfuscation. In “Proceedings of the 26th International Conference on World Wide Web”. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 627–636. DOI:https://doi.org/10.1145/3038912.3052696

[6] S. Salamatian et al., "How to hide the elephant- or the donkey- in the room: Practical privacy against statistical inference for large data," 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, 2013, pp. 269-272, DOI: 10.1109/GlobalSIP.2013.6736867.

[7] Latanya Sweeney. 2002. K-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 10, 5 (October 2002), 557–570. DOI:https://doi.org/10.1142/S0218488502001648

[8] Benjamin C. M. Fung, Ke Wang, Rui Chen, and Philip S. Yu. 2010. Privacy-preserving data publishing: A survey of recent developments. ACM Comput. Surv. 42, 4, Article 14 (June 2010), 53 pages. DOI:https://doi.org/10.1145/1749603.1749605

[9] Rakesh Agrawal and Ramakrishnan Srikant. 2000. Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD international conference on Management of data (SIGMOD '00). Association for Computing Machinery, New York, NY, USA, 439–450. DOI:https://doi.org/10.1145/342009.335438

[10] S. Banerjee, N. Hegde, and L. Massoulie, “The price of privacy in untrusted recommendation engines,” in Allerton, 2012.

[11] A. Evfimievski, J. Gehrke, and R. Srikant, “Limiting privacy breaches in privacy-preserving data mining,” in PODS, 2003

[12] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in TCC, 2006

[13] F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872