ID-SR

Privacy-Preserving Social Recommendation Based on Infinite Divisibility for Trustworthy AI

Cui, Jingyi; Xu, Guangquan; Liu, Jian; Feng, Shicheng; Wang, Jianli; Peng, Hao; Fu, Shihui; Zheng, Zhaohua; Zheng, Xi; Liu, Shaoying

DOI
10.1145/3639412

Publication date
2024

Document Version
Final published version

Published in
ACM Transactions on Knowledge Discovery from Data

ID-SR: Privacy-Preserving Social Recommendation Based on Infinite Divisibility for Trustworthy AI

JINGYI CUI, School of New Media and Communication, Tianjin University, Tianjin, China
GUANGQUAN XU, JIAN LIU, and SHICHENG FENG, College of Intelligence and Computing, Tianjin University, Tianjin, China
JIANLI WANG, School of New Media and Communication, Tianjin University, Tianjin, China
HAO PENG, Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Jinhua, China and School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China
SHIHUI FU, TU Delft, Delft, Netherlands
ZHAOHUA ZHENG, College of Intelligence and Computing, Tianjin University, Tianjin, China and School of CyberSpace Security, Hainan University, Haikou, China
XI ZHENG, School of Computing, Macquarie University, Sydney, Australia
SHAOYING LIU, School of Informatics and Data Science, Hiroshima University, Higashihiroshima, Japan

Recommendation systems powered by artificial intelligence (AI) are widely used to improve user experience. However, AI inevitably raises privacy leakage and other security issues due to the utilization of extensive user data. Addressing these challenges can protect users' personal information, benefit service providers, and foster service ecosystems. Presently, numerous techniques based on differential privacy have been proposed to solve this problem. However, existing solutions encounter issues such as inadequate data utilization and a tenuous trade-off between privacy protection and recommendation effectiveness. To enhance recommendation accuracy and protect users' private data, we propose ID-SR, a novel privacy-preserving social recommendation scheme for trustworthy AI based on the infinite divisibility of Laplace distribution. We first introduce a novel recommendation method adopted in ID-SR, which is established based on matrix factorization with a newly designed social regularization term for improving recommendation effectiveness. We then propose a

This work is supported in part by the National Science Foundation of China (grant nos. U22B2027, 62172297, 62102262, 61902276, and 62272311), the Tianjin Intelligent Manufacturing Special Fund Project (grant no. 20211097), the China Guangxi Science and Technology Plan Project—Guangxi Science and Technology Base and Talent Special Project (grant no. AD23026096; Application Number 2022AC20001), the Hainan Provincial Natural Science Foundation of China (grant no. 622RC616), and the CCF-Nsfocus Kunpeng Fund Project (grant no. CCF-NSFOCUS202207).
Authors' addresses: J. Cui and J. Wang, School of New Media and Communication, Tianjin University, Tianjin, 300110, China; e-mails: {cuijingyi, 2022245004}@tju.edu.cn; G. Xu, J. Liu (Corresponding author), and S. Feng (Corresponding author), College of Intelligence and Computing, Tianjin University, Tianjin, 300350, China; e-mails: {losin, jianliu}@tju.edu.cn, 896731676@qq.com; H. Peng, Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Jinhua, 321004, China and School of Computer Science and Technology, Zhejiang Normal University, Jinhua, 321004, China; e-mail: hpeng@zjnu.edu.cn; S. Fu, TU Delft, Delft, Netherlands; e-mail: shihui.fu@tudelft.nl; Z. Zheng, College of Intelligence and Computing, Tianjin University, Tianjin, 300350, China and School of CyberSpace Security, Hainan University, Haikou, 570228, China; e-mail: zhengzhaohua@tju.edu.cn; X. Zheng, School of Computing, Macquarie University, Sydney, Australia; e-mail: james.zheng@mq.edu.au; S. Liu, School of Informatics and Data Science, Hiroshima University, Higashihiroshima, 739-8511, Japan; e-mail: sliu@hiroshima-u.ac.jp.

ACM 1556-4681/2024/06-ART161
https://doi.org/10.1145/3639412

differential privacy-preserving scheme tailored to the above method that leverages the Laplace distribution's characteristics to safeguard user data. Theoretical analysis and experimentation evaluation on two publicly available datasets demonstrate that our scheme achieves a superior balance between privacy protection and recommendation effectiveness, ultimately delivering an enhanced user experience.

CCS Concepts: • **Security and privacy → Privacy protections; • Information systems → Recommender systems; • Computing methodologies → Artificial intelligence**;

Additional Key Words and Phrases: Social recommendation, trustworthy artificial intelligence, differential privacy, matrix factorization, Laplace mechanism

ACM Reference format:
Jingyi Cui, Guangquan Xu, Jian Liu, Shicheng Feng, Jianli Wang, Hao Peng, Shihui Fu, Zhaohua Zheng, Xi Zheng, and Shaoying Liu. 2024. ID-SR: Privacy-Preserving Social Recommendation Based on Infinite Divisibility for Trustworthy AI. *ACM Trans. Knowl. Discov. Data.* 18, 7, Article 161 (June 2024), 25 pages. https://doi.org/10.1145/3639412

1 INTRODUCTION

In the era of big data, **Artificial Intelligence (AI)** is widely used in various industries. AI-based recommender systems play a crucial role in mitigating information overload and aiding users in making efficient decisions. Designing more efficient and accurate personalized recommendation methods for target users has become a prominent research topic [9, 10, 16, 18, 19, 25, 29]. However, AI-based methods need to use a large amount of user data, which inevitably brings about security issues such as personal information leakage. Research has demonstrated that even if a recommendation system does not directly utilize users' private details, such as gender or address, a malicious attacker can still deduce sensitive information through inference, which will potentially harm the user [2]. For instance, by merely observing users' ratings of items, an attacker can infer their gender or even ascertain whether they have a specific medical condition. Illegally trafficking such sensitive personal information can lead to discrimination against victims when seeking insurance or employment, even exposing them to targeted sales pitches and scams. Accordingly, designing more trustworthy AI-based algorithms that protect user privacy while deliver high-quality recommendations is of paramount importance.

To address security concerns in AI-based recommender systems, many researchers have conducted investigations using privacy-preserving techniques such as homomorphic encryption, differential privacy, k-anonymity, and others. Nevertheless, schemes based on k-anonymity [3] restrict the assumption of the attacker's background knowledge, making it not secure enough for privacy protection. Meanwhile, numerous studies based on homomorphic encryption and other cryptographic schemes [11, 15, 31, 32] entail significant computational overhead, making them challenging to implement in practical scenarios. In contrast, differential privacy [6] offers reduced computational overhead and introduces a rigorous and quantifiable privacy concept. It constrains the magnitude of the variation in the final output of an algorithm caused by a record in the dataset. Furthermore, it ensures that an attacker, even with knowledge of all data except one, cannot infer information about the unknown data. Consequently, a recommendation method based on differential privacy serves the dual purpose of protecting user privacy and delivering high-quality recommendations.

In recent years, significant advancements have been made in recommendation schemes based on differential privacy [4, 12, 23, 38]. Zhu et al. [37] incorporated differential privacy into neighbor-based collaborative filtering algorithms by applying it to items and covariance matrices, resulting in perturbed recommendations. Friedman et al. [8] introduced the differential privacy concept

to the matrix factorization recommendation algorithm. Four different perturbation schemes: Input Perturbation, Stochastic Gradient Perturbation, Output Perturbation, and **Alternating Least Squares (ALS)** with Output Perturbation are addressed in this work. However, these schemes introduced noise and did not fully utilize the rating data. Jorgensen et al. [14] introduced a novel privacy concept called **Personalized Differential Privacy (PDP)**, which generalizes differential privacy by allowing users to specify individual privacy requirements for their data. Building upon this concept, Zhang et al. [33] developed a novel recommendation scheme called PDP-PMF, which is based on Probabilistic Matrix Factorization. Unlike traditional approaches that provide a uniform level of privacy guarantee, PDP-PMF aims to satisfy user-specified privacy requirements at the item level. This approach can yield improved recommendation results for recommender systems. By leveraging deep reinforcement learning, Xiao et al. [28] introduced an alternative user profile perturbation scheme for recommender systems. This scheme leverages differential privacy to safeguard user privacy and utilizes deep **reinforcement learning (RL)** to determine the privacy budget against inference attackers, which can increase the user privacy protection level. In addition, several other methods try to provide better privacy protection and recommendation results based on machine learning [1, 5, 22, 27, 30, 34, 35]. However, most of the above methods neglect the problem that the recommender is not fully trustworthy.

Among the schemes that consider untrustworthy servers, Hua et al. [13] introduced a scheme known as **Differentially Private Matrix Factorization (DPMF)**. This scheme employs a distributed matrix factorization model and introduces perturbations to the objective function through a trusted third party. DPMF ensures that users' ratings remain undisclosed to the server, thus, safeguarding the privacy of rating data. On this basis, Meng et al. [21] sought to tackle the challenge of achieving privacy-preserving social recommendations with personalized privacy settings. They present a new scheme, PrivSR, which enables users to model ratings and social relationships privately. It can protect users' privacy from both untrusted recommenders and friends by assigning distinct noise levels to sensitive and non-sensitive ratings. Many local differential privacy schemes have also been proposed to protect users' privacy and guard against untrustworthy servers [24, 26, 36]. However, these schemes have not yet found the optimal trade-off between privacy and recommendation effectiveness.

Generally speaking, the existing schemes still suffer from the following shortcomings.

- The current schemes suffer from inadequate utilization of information, which often simply utilize the raw rating and relationship data, ignoring the other information embedded in the dataset. This will call for a comprehensive exploration of the existing data.
- The current schemes still cannot achieve a satisfactory trade-off between privacy protection and recommendation effectiveness. The application of privacy-preserving technologies naturally affects recommendation effectiveness, exacerbated by the problem of untrustworthy servers. Some existing approaches either directly modify the original scoring data or introduce noise to the results, leading to excessive noise and impacting the quality of recommendations.

To resolve these two deficiencies, we propose ID-SR, a novel privacy-preserving social recommendation scheme for trustworthy AI based on differential privacy and matrix factorization. It aims to protect user privacy while delivering optimal recommendations, which can be seen as a solution to the AI security problem in recommender systems. Specifically, first, we introduce an improved social recommendation method, called I-SR, based on matrix factorization with a novel social regularization term. This scheme treats each item individually, accounting for the influence of social relationships and users' similar preferences on their own preferences. It can effectively utilize the information embedded in historical ratings to enhance recommendation results. Second,

by employing the Laplace differential privacy mechanism, we introduce perturbations to the objective function and divide them into smaller components based on users or items. This enables users to add noise locally, protecting the original data. Third, in the perturbation process, we leverage the infinite divisibility of the Laplace distribution and its stability to devise a noise generation scheme that aligns with I-SR. Considering the type of perturbation in the matrix decomposition section and the diverse requirements of recommender systems, we develop two generation schemes: the undifferentiated noise generation scheme and the categorical noise generation scheme. Subsequently, we provide proof of the proposed scheme's adherence to the definition of differential privacy. The features and contributions of our work are summarized as follows:

- We put forward ID-SR, a novel privacy-preserving social recommendation scheme for trustworthy AI based on the infinite divisibility of Laplace distribution. We first introduce an improved social recommendation method, I-SR, based on matrix factorization with a novel social regularization term that considers the impact of social relationships and rating similarity on user preferences. The method optimizes each item individually, allowing for the comprehensive utilization of historical ratings and social relationships to enhance recommendation quality.
- For the proposed social recommendation method, ID-SR offers a new noise generation scheme based on the infinite divisibility of the Laplace distribution and its stability to protect the user information. Considering the different requirements of the recommendation system, we introduce two objective function perturbation schemes: undifferentiated perturbation and classification optimization perturbation. In the undifferentiated perturbation scheme, all ratings are assigned equal privacy budgets. In contrast, the classification optimization scheme allocates different privacy budgets to different categories of data. These schemes address the challenge posed by untrustworthy recommendation servers and potentially malicious friend users, providing better privacy guarantees to users.
- The proposed scheme satisfies the differential privacy definition in theory. Experimental results from diverse datasets demonstrate that ID-SR can provide better recommendations while preserving user privacy, achieving a better trade-off between the two objectives.

The remainder of this article is organized as follows. Section 2 introduces the relevant preliminaries, such as differential privacy. Our scheme and the theoretical proof are proposed in Section 3. Section 4 gives the experimental results. Conclusions and future work are presented in Section 5.

2 PRELIMINARIES

2.1 Differential Privacy

Differential privacy was initially introduced by Dwork [6] in 2006. By employing a rigorous mathematical proof, differential privacy ensures that any information revealed through dataset output is perturbed to a level where individual records remain indistinguishable, preventing a third party from inferring changes, additions, or deletions to a specific record based on variations in the output. It offers the highest levels of security among the current perturbation-based privacy protection methods. Formally, the differential privacy is defined as follows.

Definition 2.1 (ε-differential Privacy) Given two neighboring datasets D, D' and a randomized algorithm A, A satisfies ε-differential privacy if for any anonymized output $O \in$ Range(A),

$$Pr[A(D) = O] \leq e^{\varepsilon} \times Pr[A(D') = O],$$

where Range(A) denotes the output range of algorithm A and the term "neighboring datasets" refers to two datasets that differ by only one record. This can occur in two scenarios: either D' has

one additional or less data compared to D or D and D' have the same number of data, but only one of them differs in content. Pr denotes the probability distribution and privacy budget ε is a positive real number. The smaller it is, the better privacy protection it can provide.

This definition indicates that algorithm A satisfies ε-differential privacy if the outputs of algorithm A on any two neighboring databases are indistinguishable, i.e., the probability distributions of the outputs are less different than ε.

2.2 Laplace Mechanism

The Laplace mechanism is a widely employed differential privacy protection mechanism for numerical data. It introduces random perturbation noise to the data in order to ensure privacy protection. To illustrate the specific principles of the Laplace mechanism, we first present the definitions of sensitivity and Laplace distribution.

Definition 2.2 (Sensitivity). The maximum value of the variation of an algorithm A over two neighboring datasets D and D' is known as sensitivity, denoted as

$$\triangle A = \max_{D,D'} ||A(D) - A(D')||_1,$$

where $|| \cdot ||_1$ denotes the L_1 norm.

Definition 2.3 (Laplace Distribution [17]). The Laplace distribution $L(\mu, b)$ is a continuous probability distribution characterized by a location parameter (μ) and a scale parameter (b). The mathematical expectation of the Laplace distribution is equal to μ, the variance is equal to $2b^2$, and its probability density function is represented as follows:

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}.$$

The Laplace mechanism adds noise Y that conforms to the Laplace distribution to the result of the randomized algorithm A as follows:

$$M(D) = A(D) + Y,$$

where M(D) represents the confusion result after the addition of noise and $Y \sim L(0, \triangle A/\varepsilon)$, $\triangle A$ is the sensitivity of the algorithm A described above.

A more detailed description of the Laplace mechanism can be found in Dwork et al. [7]. We give just a brief introduction here.

2.3 Properties of the Laplace Distribution

2.3.1 *Infinite Divisibility of the Laplace Distribution.*

Definition 2.4 (Infinite Divisibility). A probability distribution with characteristic function ψ is infinitely divisible if, for any integer $n \geq 1$, we have that $\psi = \phi_n^n$, where ϕ_n is another characteristic function. In other words, a random variable Y with characteristic function ψ has the representation

$$Y \stackrel{d}{=} \sum_{i=1}^{n} X_i \tag{1}$$

for some individually and identically distributed random variables X_i.

Proposition 2.5. *Let $Y \sim L(\theta, s)$ have a Laplace distribution with characteristic function $\psi_Y(t) = e^{it\theta}/1 + s^2t^2$, $-\infty < t < +\infty$. Then, the distribution of Y is infinitely divisible. Furthermore, for every*

integer $n \geq 1$, representation (1) holds. Each X_i is distributed as $\theta/n + Y_{1n} - Y_{2n}$, where Y_{1n} and Y_{1n} are individually and identically distributed with gamma density

$$\frac{(1/s)^{1/n}}{\Gamma(1/n)} x^{\frac{1}{n}-1} e^{-x/s}, x \geq 0,$$

in which $\Gamma(1/n)$ is a gamma function. For all positive integers, $\Gamma(n) = (n-1)!$

The infinite divisibility of the classical Laplace distribution (where θ equals 0) is also stable. Thus, the following equation holds:

$$Y \stackrel{d}{=} \sqrt{B_{n-1}}(Y_1 + \cdots + Y_n),$$

where Y, $Y_i \sim L(0, s)$. B_{n-1} is a random variable drawn from a beta distribution with parameters $(n-1)$ and 1.

2.3.2 Mixture of Normal Distributions.

Proposition 2.6. *A standard classical Laplace random variable Y has the representation*

$$Y \stackrel{d}{=} \sqrt{2W}Z,$$

where the random variables W and Z have the standard exponential and normal distributions, respectively.

2.4 Matrix Factorization

Matrix factorization (MF) is a popular used AI technique in recommendation systems [20]. Its key idea is to factorize the high-dimensional and sparse rating matrix $\mathbf{R}$ into two low-dimensional and dense matrices, i.e., the user profile matrix $\mathbf{U}$ and the item profile matrix $\mathbf{V}$. Then, the two profile matrices will be used to predict user ratings of unrated items in the item set.

Assuming that there are n users and m items in the recommender system, we use $\mathbf{R}_{ij}$ to denote the rating of user i on item j. The original rating matrix $\mathbf{R} = [\mathbf{R}_{ij}] \in \mathbb{R}^{n \times m}$ is typically sparse since, in general, users rate a much smaller number of items compared with the total number of items available. MF decomposes $\mathbf{R}$ into two matrices, $\mathbf{U} = [\mathbf{u}_i]_{i \in [n]} \in \mathbb{R}^{d \times n}$ and $\mathbf{V} = [\mathbf{v}_j]_{j \in [m]} \in \mathbb{R}^{d \times m}$. Here, $\mathbf{u}_i$ represents the latent vector of user i, $\mathbf{v}_j$ represents the latent vector of item j, and d denotes the number of latent factors, which can react to the implicit characteristics of users and items. For instance, in a user movie rating dataset, the hidden factors of items may represent different movie genres, such as action, comedy, and more. The latent vectors of a specific movie can be interpreted as its degree of association with various genres. To enhance the recommendation effectiveness, various variants of matrix factorization techniques exist. The more commonly objective function of MF with regularization terms is as follows:

$$\min_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{I}_{ij} \left(\mathbf{R}_{ij} - \mathbf{u}_i^T \mathbf{v}_j \right)^2 + \lambda \left(||\mathbf{u}_i||^2 + ||\mathbf{v}_j||^2 \right),$$

where $\mathbf{I}_{ij} = 1$ if user i rated item j; otherwise, $\mathbf{I}_{ij} = 0$.

The final profile matrix $\mathbf{U}$ for users and $\mathbf{V}$ for items can be obtained using the stochastic gradient descent algorithm to optimize the function above iteratively. The inner product of these two matrices results in $\hat{\mathbf{R}} = \mathbf{U} \times \mathbf{V}$, which can be utilized for rating prediction.

2.5 Social Recommendation Based on Matrix Factorization

Incorporating social information into matrix factorization has resulted in various schemes. In this study, we specifically examine the recommendation method proposed by Ma et al. [19] that utilizes

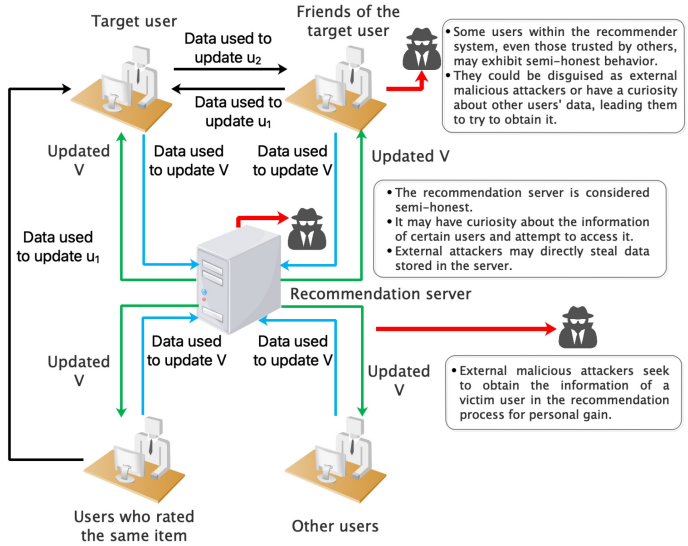


Fig. 1. The framework of ID-SR.

social regularization terms. Based on the above MF algorithm, the social recommendation method can be mathematically written as follows:

$$\min_{\mathrm{U,V}} \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{I}_{ij}\left(\mathbf{R}_{ij}-\mathbf{u}_i^T\mathbf{v}_j\right)^2+\lambda\left(||\mathbf{u}_i||^2+||\mathbf{v}_j||^2\right)+\sum_{i=1}^{n}\sum_{f\in F_i} S_{if}||\mathbf{u}_i-\mathbf{u}_f||_F^2, \tag{2}$$

where F_i denotes the friends of user i, S_{if} denotes the similarity between user i and user f, and $||\cdot||_F^2$ denotes the Frobenius norm.

3 THE PROPOSED SOCIAL RECOMMENDATION WITH DIFFERENTIAL PRIVACY

3.1 Scenario

In this work, we consider the scenario in which the recommendation server is not entirely trustworthy. We aim to prevent the disclosure of the accurate contents of the original data through the perturbation of the item and user latent vector. By doing so, we safeguard the privacy of each user's information within the system, protecting it from potential inference. The framework of our scheme, depicted in Figure 1, involves two entities: the recommendation server and the users. The recommendation server assumes the responsibility of maintaining the item profile matrix, aggregating the perturbed data from each user and updating the latent vectors of each item to complete the item profile matrix. Users are required to manage their own latent vectors and transmit the perturbed data to both the recommendation server and other users to update the latent vectors.

For clarity, we will primarily focus on describing the recommendation scheme from the perspective of one user. To this end, we divide the user entity into four categories:

—The target user, who is the recipient of recommendations
—The friends of the target user, who are the users that have a social relationship with the target user.
—The users who rated the same items as the target user
—Other users who do not belong to the above entities

It is important to note that all users in the four categories are actual participants in the recommendation system and will be included as the recommended target user.

3.2 Threat Model

In practical scenarios, users, recommendation servers, and external malicious attackers may attempt to access the personal information of specific users due to economic interests and other reasons. Therefore, we consider the following threat models:

- —An external malicious attacker seeks to obtain the data of a victim user in the recommendation process and subsequently steal the victim's private information for personal gain.
- —The recommendation server is considered semi-honest, strictly following the scheme to provide recommendation results, but may be curious about certain users' accurate data and attempt to access it. At the same time, even if the server is trusted, an attacker may be able to directly steal the data stored in it in order to gain access to the user's private information.
- —Some users within the recommendation system, even those trusted by others, may exhibit semi-honest behavior. External attackers could disguise themselves as normal users or some normal users may be curious about other users' data, leading to the semi-honest behavior.

3.3 Settings

To address all the security threats mentioned above and simultaneously harness the full potential of the data for improved recommendation results, our scheme, ID-SR, is structured as follows. First, we maximize the utilization of information within the rating data by introducing a novel social regularization term. Building upon this, we propose an improved MF-based social recommendation method (I-SR). This method comprehensively accounts for the impact of social relationships and rating similarity on user preferences, adopts individualized optimization for each item instead of uniform treatment, and thus employs a fine-grained optimization scheme for the item set, resulting in improved recommendation effectiveness. Second, to ensure privacy protection for user data, we devise an associated perturbation scheme for I-SR. It is based on the Laplace differential privacy mechanism and leverages the infinite divisibility of the Laplace distribution. Within the MF section, we introduce two types of perturbation schemes: the undifferentiated scheme and the classification optimization scheme. The distinction between these two schemes is whether they categorize and incorporate the rating data with varying privacy levels. They are suitable for different systems.

In our scheme, the item profile matrix $\mathbf{V}$, perturbed by each user, is maintained by the recommendation server and will eventually be disclosed to the user to provide recommendations. Since the recommendation server is considered semi-honest, the user's own latent vector $\mathbf{u}_i$ should be maintained locally and secretly by oneself to prevent the attackers from deducing the user's rating of an item by calculating $\mathbf{u}_i \times \mathbf{V}$. The communication data between the user and the server or between the users are perturbed during the whole scheme process. Therefore, a privacy guarantee can be provided. To facilitate subsequent discussions, we summarize the critical notations employed in Table 1.

3.4 Improved Social Recommendation

Social recommendation uses historical rating data and social relationship information of users to provide personalized recommendations. Fully utilizing the information contained in these data can provide better recommendations. In this work, we begin with proposing an improved social recommendation method, I-SR, which features a novel social regularization term.

Table 1. Critical Notations Employed in ID-SR

Notation	Meaning
n	Number of users
m	Number of items
d	Number of latent factors
$\mathbf{u}_i$	Latent vector of user i
$\mathbf{v}_j$	Latent vector of item i
$\mathbf{o}_j^i$	User i's noise vector for item j
$\mathbf{U}$	User profile matrix, composed by $\mathbf{u}_i$
$\mathbf{V}$	Item profile matrix, composed by $\mathbf{v}_j$
$\mathbf{R}_{ij}$	Rating of item j by user i
$\mathbf{I}_{ij}$	Indicator function, whether user i rated item j or not
R_{max}	Maximum rating value in the system
R_{min}	Minimum rating value in the system
S_{if}	Similarity between user i and user f
S_{ij}^f	For item j, the similarity between user i and user f
C_j	Users who rated item j in addition to the target user
F_i	Friends of user i
$Exp(\lambda)$	Exponential distribution with parameter λ
B_n	Beta distribution with parameter n and 1
$L(b)$	Laplace distribution with position parameter 0 and scale parameter b
$\Gamma(k, \theta)$	Gamma distribution with shape parameter k and scale parameter θ
$N(\mu, \sigma^2)$	Normal distribution with location parameter μ and squared scale parameter σ^2
$U(a, b)$	Continuous uniform distribution with parameters a and b

We consider a recommender system consisting of n users and m items. The system includes the rating records $\mathbf{R} = [\mathbf{R}_{ij}] \in \mathbb{R}^{n \times m}$ as well as the social relationships between users. In this work, we solely focus on directed binary social relationships. Specifically, a value of 1 indicates when user i trusts user f, whereas 0 indicates no trust. Typically, the recommendation method incorporating social regularization is denoted as Equation (2). The method consists of two components: the basic matrix factorization model and the social regularization term, in which the regularization term captures the influence of social relationships on user preferences. However, we identified two disadvantages with this term. First, it solely focuses on the influence of social relationships, specifically users' trusted friends, on the target user's preferences, which fails to leverage the rich information present in the historical rating data. In fact, the historical rating data contains the rated records of many users. Based on the idea that users with similar preferences may prefer a specific target item, we propose that the rating data can be more fully exploited to optimize the regularization term by adding $\sum_{i=1}^{n} \sum_{c \in C_j} S_{ic} ||\mathbf{u}_i - \mathbf{u}_c||_F^2$ as a complementary, where C_j denotes the users who rated item j in addition to the target user and S_{ic} indicates the similarity between the two users, which is usually computed by **Vector Space Similarity (VSS)** and the **Pearson Correlation Coefficient (PCC)** [19]. This term shows the effect of users with similar preferences on the target user, which can also be understood as a potential social relationship and coincides with the idea of user-based collaborative filtering.

The second question is this: do users with social relationships who share a preference for one item necessarily share preference for another thing? The answer is no, as gender, age, and only individual preferences significantly impact user preferences. Current social regularization items

are user based, assigning equal weight to each item rated by a user. Instead, we propose computing the inter-user similarity for each item at a fine-grained level, that is, changing the traditional term to $\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{f\in F_i} S_{ij}^{f}(||\mathbf{u}_i-\mathbf{u}_f||)_F^2$, where S_{ij}^{f} denotes the similarity between users i and f for item j. The existing similarity measures are based on multiple vectors and do not apply to our proposed new similarity. Therefore, in I-SR, we use the following formula to measure the similarity between two users for item j:

$$S_{ij}^{f}=1-\frac{|\mathbf{R}_{ij}-\mathbf{R}_{fj}|}{R_{max}-R_{min}}, \tag{3}$$

where R_{max} and R_{min} represent the maximum and minimum values of rating allowed in the system, respectively.

By addressing the two issues above, I-SR is able to provide an accurate measurement of the impact of social relationships and preference similarity on the target user at the item level. In summary, the objective function of our improved social recommendation method, based on the basic matrix factorization model, is as follows:

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} J=&\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{I}_{ij}\left(\mathbf{R}_{ij}-\mathbf{u}_i^T\mathbf{v}_j\right)^2+\lambda\left(||\mathbf{U}||_F^2+||\mathbf{V}||_F^2\right)\\&+\alpha\sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{f\in F_i}S_{ij}^{f}(||\mathbf{u}_i-\mathbf{u}_f||)_F^2+\alpha\sum_{i=1}^{n}\sum_{j=1}^{|V_i|}\sum_{c\in C_j}S_{ij}^{c}(||\mathbf{u}_i-\mathbf{u}_c||)_F^2,\end{aligned} \tag{4}$$

in which λ denotes the coefficients of the matrix regularization term, α indicates the coefficients of the social regularization terms, and $|V_i|$ represents the number of items user i has rated. Like traditional methods, I-SR consists of two parts: the basic matrix factorization model and the social regularization term. Evaluations on two publicly available datasets show that by extending and accurately carving the regularization terms, I-SR can provide better recommendations, which will be given in Section 4.

3.5 Perturbation Scheme Based on Differential Privacy

In order to protect sensitive information, such as users' ratings and latent vectors $\mathbf{u}_i$, we employ differential privacy in our proposed scheme. More specifically, we utilize objective function perturbation to safeguard the confidentiality of the generated item and user profile matrices to protect users' information. Considering the threat model outlined in Section 3.2, we adopt the strategy of individual optimization of $\mathbf{U}$ and $\mathbf{V}$, rather than joint optimization, to mitigate the risk of multiple attackers gaining access to the information. Within ID-SR, we employ the gradient descent method to update the item and user latent vectors. Guided by Equation (4), the gradients with respect to $\mathbf{u}_i$ and $\mathbf{v}_j$ are as follows:

$$\frac{\partial J}{\partial \mathbf{v}_j}=2\sum_{i=1}^{n}\mathbf{I}_{ij}\left(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij}\right)\mathbf{u}_i+2\lambda\mathbf{v}_j \tag{5}$$

$$\frac{\partial J}{\partial \mathbf{u}_i}=2\sum_{j=1}^{m}\mathbf{I}_{ij}\left(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij}\right)\mathbf{v}_j+2\lambda\mathbf{u}_i+\alpha\sum_{j=1}^{|V_i|}\sum_{f\in F_i}2S_{ij}^{f}(\mathbf{u}_i-\mathbf{u}_f)+\alpha\sum_{j=1}^{|V_i|}\sum_{c\in C_j}2S_{ij}^{c}(\mathbf{u}_i-\mathbf{u}_c), \tag{6}$$

where F_i stands for the friends of user i.

Among them, $\mathbf{V}$ is publicly available to provide user recommendation services and simultaneously shared with other recommendation systems for joint optimization. The user's latent vector, $\mathbf{u}_i$, and the specific rating value, $\mathbf{R}_{ij}$, are sensitive information that should only be possessed by the

user.[1] However, the gradient concerning $\mathbf{v}_j$ as Equation (5), which is employed by the recommendation server to update $\mathbf{v}_j$, unavoidably relies on this sensitive information. Hence, it is crucial to introduce perturbation in the basic matrix factorization model to prevent the recommendation server and the external attackers from extracting accurate data through inference attacks and other methods. Moreover, based on Equation (6), it is evident that when users update their own hidden vectors, $\mathbf{u}_i$, they rely on the vectors, $\mathbf{u}_f$ and $\mathbf{u}_c$, which are from their friends or other users who have rated the same item. If unprotected, the information of user f and c may be leaked. Consequently, it is essential to introduce perturbation in the social regularization term to protect against user-based and external attackers. Subsequently, we will present the perturbation schemes for both the matrix factorization model and the social regularization term.

3.5.1 Privacy-Preserving Matrix Factorization. Based on the above discussion, each user who rated j needs to send $2\mathbf{I}_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{ij})\mathbf{u}_i$ to the recommendation server for updating $\mathbf{v}_j$. To protect the sensitive information contained therein, we apply the objective perturbation method [4] with ε-differential privacy. First, we introduce a formal objective function of the matrix factorization model with perturbations:

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{I}_{ij}\left(\left(\mathbf{R}_{ij}-\mathbf{u}_i^T\mathbf{v}_j\right)^2+\mathbf{v}_j^T\mathbf{o}_j^i\right)+\lambda\left(||\mathbf{U}||_F^2+||\mathbf{V}||_F^2\right), \tag{7}$$

where $\mathbf{o}_j^i \in \mathbb{R}^{d\times 1}$ is user i's noise vector for item j. The gradient with respect to $\mathbf{v}_j$ according to Equation (7) is as follows:

$$\frac{\partial J}{\partial \mathbf{v}_j}=\sum_{i=1}^{n}\mathbf{I}_{ij}\left(2\left(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij}\right)\mathbf{u}_i+\mathbf{o}_j^i\right)+2\lambda\mathbf{v}_j. \tag{8}$$

At this point, users can send the perturbed gradient data $\mathbf{I}_{ij}(2(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij})\mathbf{u}_i+\mathbf{o}_j^i)$ to the server, which will effectively hide the user's sensitive information and prevent it from being stolen. In order to provide $\varepsilon-$differential privacy guarantees for $\mathbf{v}_j$, $\sum_{i=1}^{n}\mathbf{o}_j^i=\mathbf{o}_j\sim L(2\triangle\sqrt{d}/\varepsilon)$ is required. However, ensuring that the sum of each noise conforms to the Laplace distribution is challenging if users are allowed to generate noise arbitrarily. Thus, Hua et al. [13] proposed a solution based on Proposition 2.6 involving interactions between the server and users. In this work, we propose a solution based on the infinite divisibility of the Laplace distribution, employing it as **the matrix factorization part of our undifferentiated scheme**.

First, each user who rated item j randomly selects the noise vector $Y_j^{1i}, Y_j^{2i}\in\mathbb{R}^{d\times 1}$, where each element of them is randomly and independently picked from $\Gamma(1/|R_j|, 2\triangle\sqrt{d}/\varepsilon)$. $|R_j|$ denotes the number of users who rated item j. Then, the user just computes $\mathbf{o}_j^i=Y_j^{1i}-Y_j^{2i}$ as the final noise vector. According to Proposition 2.5, we have that $\mathbf{o}_j\sim L(2\triangle\sqrt{d}/\varepsilon)$. Our perturbation scheme is generated locally by the users, which reduces the interaction requirement compared with the existing scheme.

Theorem 3.1. *Let $\triangle=R_{max}-R_{min}$. If each element in $\mathbf{o}_j$ is randomly and independently selected from $L(2\triangle\sqrt{d}/\varepsilon)$, the derived V satisfies $\varepsilon-$ differential privacy.*

Proof. Based on Proposition 2.5, if each element of Y_j^{1i}, Y_j^{2i} is randomly and independently selected from $\Gamma(1/|R_j|, 2\triangle\sqrt{d}/\varepsilon)$, that is, with the characteristic function

$$(1-2\triangle it\sqrt{d})^{-1/|R_j|},$$

[1]If an attacker gets hold of $\mathbf{u}_i$ or gets an approximation of $\mathbf{u}_i$ through inference attack, he can compute $\mathbf{u}_i\times\mathbf{V}$ to obtain the user's ratings for all items and then use these to infer personal information about the victim, such as the health status.

where i is an imaginary unit and t is a real number, then, by the nature of the characteristic function, the characteristic function of each element in $o_j^i = Y_j^{1i} - Y_j^{2i}$ is

$$\phi = \left(\frac{1}{1-(2\triangle\sqrt{d}/\varepsilon)it}\right)^{1/|R_j|} \times \left(\frac{1}{1+(2\triangle\sqrt{d}/\varepsilon)it}\right)^{1/|R_j|} = \left(\frac{1}{1-(2\triangle\sqrt{d}/\varepsilon)^2 i^2 t^2}\right)^{1/|R_j|} = \left(\frac{1}{1+(2\triangle\sqrt{d}/\varepsilon)^2 t^2}\right)^{1/|R_j|}.$$

Thus, we have that

$$\phi^{|R_j|} = \frac{1}{1+(2\triangle\sqrt{d}/\varepsilon)^2 t^2} = \psi,$$

in which ψ is the characteristic function of $L(0, 2\triangle\sqrt{d}/\varepsilon)$. In other words, we have that $\sum_{i=1}^{|R_j|} \mathbf{o}_j^i = \mathbf{o}_j \sim L(0, 2\triangle\sqrt{d}/\varepsilon)$, which means that each coordinate o_{jl} of $\mathbf{o}_j = (o_{j1}, o_{j2}, \ldots, o_{jd})$ is from the $L(\frac{2\triangle\sqrt{d}}{\varepsilon})$. The density function of it is $\frac{\varepsilon}{4\triangle\sqrt{d}} e^{-\frac{\varepsilon|o_{jl}|}{2\triangle\sqrt{d}}}$.

Let D, D' be two neighboring datasets only differing from one record $\mathbf{R}_{ij}$ and $\tilde{\mathbf{R}_{ij}}$ since, from the different inputs, we obtain the same output V, which is obtained after convergence. According to Equation (8), we have that $\frac{\partial J(D)}{\partial \mathbf{v}_j} = \frac{\partial J(D')}{\partial \mathbf{v}_j} = 0$, that is,

$$2\sum_{i=1}^{n} \mathbf{I}_{ij}\left(\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{ij}\right)\mathbf{u}_i + \mathbf{o}_j = 2\sum_{i=1}^{n} \mathbf{I}_{ij}\left(\mathbf{u}_i^T\mathbf{v}_j - \tilde{\mathbf{R}_{ij}}\right)\mathbf{u}_i + \tilde{\mathbf{o}_j}.$$

Since only the records $\mathbf{R}_{ij}$ and $\tilde{\mathbf{R}_{ij}}$ make a difference, we get that

$$\mathbf{o}_j - \tilde{\mathbf{o}_j} = 2\mathbf{u}_i(\mathbf{R}_{ij} - \tilde{\mathbf{R}_{ij}}).$$

Considering that $|\mathbf{R}_{ij} - \tilde{\mathbf{R}_{ij}}| \le \triangle$ and $||\mathbf{u}_i|| \le 1$, we have that

$$||\mathbf{o}_j - \tilde{\mathbf{o}_j}|| \le 2\triangle.$$

For each vector $\mathbf{v}_j$ of the derived $\mathbf{V}$, we can get that

$$\frac{Pr[\mathbf{v}_j|D]}{Pr[\mathbf{v}_j|D']} = \frac{\prod_{l=1}^{d} Pr(o_{jl})}{\prod_{l=1}^{d} Pr(\tilde{o_{jl}})} = e^{-\frac{\varepsilon\sum_{l=1}^{d}|o_{jl}|}{2\triangle\sqrt{d}}} / e^{-\frac{\varepsilon\sum_{l=1}^{d}|\tilde{o_{jl}}|}{2\triangle\sqrt{d}}} = e^{\frac{\varepsilon\sum_{l=1}^{d}(|o_{jl}|-|\tilde{o_{jl}}|)}{2\triangle\sqrt{d}}} \le e^{\frac{\varepsilon\sqrt{d\sum_{l=1}^{d}(|o_{jl}|-|\tilde{o_{jl}}|)^2}}{2\triangle\sqrt{d}}} = e^{\frac{\varepsilon\sqrt{d}||o_{jl}-\tilde{o_{jl}}||}{2\triangle\sqrt{d}}} \le e^{\varepsilon}.$$

Thus, we can provide the privacy guarantee consistent with $\varepsilon-$differential privacy for the derived $\mathbf{V}$. □

The approach mentioned above treats all historical rating data uniformly. In contrast, many contemporary recommendation systems offer users the capability to categorize ratings or other shared content. Providing unequal weights of privacy protection for different types of ratings can significantly enhance the recommendation quality of the recommendation system within the same privacy budget. Hence, following the concept of personalized social recommendation proposed by Meng et al. [21], we divide the types of ratings and employ diverse levels of privacy protection as **the matrix factorization part of our classification optimization scheme**. In the classification optimization scheme, the objective function of the matrix decomposition part becomes

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} J = &\sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{I}_{ij}^1\left(\left(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j\right)^2 + \mathbf{v}_j^T\mathbf{o}_{1,j}^i\right) + \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{I}_{ij}^2\left(\left(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j\right)^2 + \mathbf{v}_j^T\mathbf{o}_{2,j}^i\right) + \cdots \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{I}_{ij}^K\left(\left(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j\right)^2 + \mathbf{v}_j^T\mathbf{o}_{K,j}^i\right) + \lambda\left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\right),\end{aligned}$$

where $\mathbf{I}_{ij}^k (k = 1, 2, \ldots, K)$ is the indicator function for ratings of category k, which denotes whether user i rated item j or not. $\mathbf{o}_{k,j}^i$ is the noise generated by user i for item j from each category. In our

scheme, the noise size is different for different categories, and private data requires a larger noise than fully publicizable data to provide more privacy protection.

Currently, different recommendation systems set up many different sharing categories for the content sharing session, for example, setting the sharing content into two categories: private and public or providing more categories of categorization. To facilitate descriptive convenience, we will illustrate our noise generation scheme using a 3-categorized (private, partially public, and fully public) recommender system as an example. Assume that each of the three types is assigned an ε_1, ε_2, and ε_3 privacy budget, where $\varepsilon_1 = \beta_1\varepsilon_3, \beta \in (0,1], \varepsilon_2 = \beta_2\varepsilon_3, \beta_2 \in (0,1]$, and $\beta_1 \le \beta_2$. In the scenario in which the total privacy budget is ε, we set $\varepsilon_3 = (\frac{1}{\beta_1} + \frac{1}{\beta_2} + 1)\varepsilon$ (a similar setup could be adopted for more categories) and employ the interactive noise design for generating noise based on the additivity of the normal distribution. Specifically, according to Proposition 2.6, the recommendation server first generates $\mathbf{W}_j \in \mathbb{R}^{d\times 1}$ for item j, where each element of it is randomly and independently selected from $Exp(1)$. Then, the server shares $\mathbf{W}_j$ to the users in $R_{3,j}$ who rated the item j and these users select $\mathbf{Z}_j^i \in \mathbb{R}^{d\times 1}$. Here, there is a difference in the $\mathbf{Z}_j^i$ generated from users who categorize the ratings of this item. For the users who set the rating fully public, each element in $\mathbf{Z}_j^i$ (denote as $\mathbf{Z}_{3,j}^i$) is randomly and independently selected from $N(0, 1/|R_{3,j}|)$ (similar for the users who set the rating to other categories), where $|R_{3,j}|$ denotes the number of users who set the rating of j as fully public. Thus, the noise vector $\mathbf{o}_{3,j}^i$ can be computed as $\mathbf{o}_{3,j}^i = 2 \triangle (2d\mathbf{W}_j)^{1/2}\mathbf{Z}_{3,j}^i/\varepsilon_3$. Based on the additivity of the normal function, we have that $\sum_{i=1}^{|R_{n,j}|} \mathbf{o}_{3,j}^i = \mathbf{o}_{3,j} \sim L(2 \triangle \sqrt{d}/\varepsilon_3)$, which is equivalent to each element in $\mathbf{o}_{3,j}$ being independently and randomly selected from $L(2 \triangle \sqrt{d}/\varepsilon_3)$.

Theorem 3.2. *Let $\triangle = R_{max} - R_{min}$. If each element in $\mathbf{o}_{1,j}, \mathbf{o}_{2,j}, \ldots, \mathbf{o}_{K,j}$ is randomly and independently selected from $L(2 \triangle \sqrt{d}/\varepsilon_1)$, $L(2 \triangle \sqrt{d}/\varepsilon_2), \ldots, L(2 \triangle \sqrt{d}/\varepsilon_K)$, respectively, then the derived V satisfies $\varepsilon-$differential privacy.*

Proof. Since $\varepsilon_K = (\frac{1}{\beta_1} + \frac{1}{\beta_2} + \cdots + \frac{1}{\beta_{K-1}})\varepsilon$ and $\varepsilon_j = \beta_j\varepsilon_K$, for $j = 1, \ldots, K-1$, we can formulate the summation of all the random noise from all the users as follows:

$$\begin{aligned}\mathbf{o}_j &= \sum_{i=1}^{|R_{1,j}|} \mathbf{o}_{1,j}^i + \sum_{i=1}^{|R_{2,j}|} \mathbf{o}_{2,j}^i + \cdots + \sum_{i=1}^{|R_{K,j}|} \mathbf{o}_{K,j}^i \\ &= \frac{2 \triangle \sqrt{2d\mathbf{W}_j}}{\varepsilon_1} \sum_{i=1}^{|R_{1,j}|} \mathbf{Z}_{1,j}^i + \frac{2 \triangle \sqrt{2d\mathbf{W}_j}}{\varepsilon_2} \sum_{i=1}^{|R_{2,j}|} \mathbf{Z}_{2,j}^i + \cdots + \frac{2 \triangle \sqrt{2d\mathbf{W}_j}}{\varepsilon_K} \sum_{i=1}^{|R_{K,j}|} \mathbf{Z}_{K,j}^i.\end{aligned}$$

Based on the additivity of the normal distribution (can be denoted as $\sum_{i=1}^{n} N(0, \frac{1}{n}) = N(0,1)$), the above equation can be expressed as

$$\begin{aligned}\mathbf{o}_j &= 2 \triangle \sqrt{2d\mathbf{W}_j}\mathbf{Z}_j \left(\frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} + \cdots + \frac{1}{\varepsilon_K}\right) \\ &= 2 \triangle \sqrt{2d\mathbf{W}_j}\mathbf{Z}_j \left(\frac{1}{(1 + \frac{\beta_1}{\beta_2} + \cdots + \frac{\beta_1}{\beta_{K-1}})\varepsilon} + \frac{1}{(\frac{\beta_2}{\beta_1} + 1 + \cdots + \frac{\beta_2}{\beta_{K-1}})\varepsilon} + \cdots + \frac{1}{(\frac{1}{\beta_1} + \frac{1}{\beta_2} + \cdots + \frac{1}{\beta_{K-1}})\varepsilon}\right) \\ &= \frac{2 \triangle \sqrt{d}}{\varepsilon}\sqrt{2\mathbf{W}_j}\mathbf{Z}_j.\end{aligned}$$

where $\mathbf{Z}_j$ denotes the standard normal distribution that can be indicated as $\mathbf{Z}_j \sim N(0,1)$. Thus, we can get that the summation of all the random noise from all the users, $\mathbf{o}_j = (o_{j1}, o_{j2}, \ldots, o_{jd})$, is a d-dimensional vector, in which each coordinate $o_{jl} \sim L(2 \triangle \sqrt{d}/\varepsilon)$.

The subsequent proof is identical to the proof of Theorem 3.1 for V satisfying $\varepsilon-$differential privacy; our classification optimization scheme can provide the same privacy guarantee for V. □

3.5.2 Privacy-Preserving Social Regularization. During the process of updating the user latent vector $\mathbf{u}_i$, it is necessary for each user to transmit the data, including one's own latent vector, ratings, and other relevant information for other users to update $\mathbf{u}_i$. To safeguard the privacy of this data, we also employ objective function perturbation. The traditional perturbation scheme of the social regularization term can be expressed as follows:

$$\min_{\mathrm{U,V}} \sum_{i=1}^{n} \sum_{f \in F_i} \left(S_{if} ||\mathbf{u}_i - \mathbf{u}_f||_F^2 + \mathbf{u}_i^T \mathbf{o}_i^f \right).$$

In ID-SR, the incorporation of the novel social regularization term necessitates the adoption of a new perturbation design in order to ensure $\varepsilon-$differential privacy protection. Leveraging the stability properties of the Laplace distribution, we propose the following perturbation scheme:

$$\min_{\mathrm{U,V}} \sum_{i=1}^{n} \sum_{j=1}^{|V_i|} \sum_{f \in F_i} \left(S_{ij}^f ||\mathbf{u}_i - \mathbf{u}_f||_F^2 + \mathbf{u}_i^T \sqrt{B_1^i} \mathbf{o}_{ij}^f \right) + \sum_{i=1}^{n} \sum_{j=1}^{|V_i|} \sum_{c \in C_j} \left(S_{ij}^c ||\mathbf{u}_i - \mathbf{u}_c||_F^2 + \mathbf{u}_i^T \sqrt{B_1^i} \mathbf{o}_{ij}^c \right), \quad (9)$$

where $\sqrt{B_1^i} \in \mathbb{R}^{d \times 1}$ is a vector generated by the target user and each element in it is independently and randomly selected from a Beta distribution with parameters both being 1. $\mathbf{o}_{ij}^f$ and $\mathbf{o}_{ij}^c$ are the noise vectors generated by other users for the target user i.

In I-SR, during the process of calculating the similarity between two users (S_{ij}^f and S_{ij}^c), private data $\mathbf{R}_{ij}$ from the other users are used; it also needs to be processed to protect the users' information. Thus, in ID-SR, we change the similarity as follows:

$$S_{ij}^f = 1 - \frac{|\mathbf{R}_{ij} + q_{ij}^f - \mathbf{R}_{fj}|}{R_{max} - R_{min}},$$

where $q_{ij}^f \sim U(R_{min}, R_{max})$ is a random variable selected from the uniform distribution. At the same time, to keep the meaning of similarity, we make the following settings: if the perturbed S_{ij}^f exceeds 1, then set it to 1, whereas if it is less than 0, then set it to 0.

Similarly, in order to provide a $\varepsilon-$differential privacy guarantee for each derived $\mathbf{u}_i$, we need to design a noise generation scheme to make $\sum_{j=1}^{|V_i|} \sum_{f \in F_i} \mathbf{o}_{ij}^f = \mathbf{o}_{f,i} \sim L(4\sqrt{d}/\varepsilon)$ and $\sum_{j=1}^{|V_i|} \sum_{c \in C_j} \mathbf{o}_{ij}^c = \mathbf{o}_{c,i} \sim L(4\sqrt{d}/\varepsilon)$. Both of the methods we introduced in Section 3.5.1 can be used. Specifically, in our proposed approach based on the infinite divisibility of the Laplace distribution, the users in F_i (similar to the users in C_j) generate the vector $\mathbf{o}_{ij}^f = Y_i^{1f} - Y_i^{2f}$, where each element of $Y_i^{1f} - Y_i^{2f}$ is randomly and independently picked from $\Gamma(1/(|V_i| * |F_i^+|), 4\sqrt{d}/\varepsilon)$. In the interactive generation scheme, the target user first generates $\mathbf{W}_i \sim Exp(1)$ and the users in F_i^+ select $\mathbf{Z}_i^f \sim N(0, 1/(|V_i| * |F_i^+|))$; the noise vector can be computed as $\mathbf{o}_{ij}^f = 4(2d\mathbf{W}_i)^{1/2}\mathbf{Z}_i^f/\varepsilon$. According to our objective function Equation (9), the gradient of the socialization part with respect to $\mathbf{u}_i$ is as follows:

$$\frac{\partial J}{\partial u_i} = \sum_{j=1}^{m} \sum_{f \in F_i} (2S_{ij}^f(\mathbf{u}_i - \mathbf{u}_f) + \sqrt{B_1^i}\mathbf{o}_{ij}^f) + \sum_{j=1}^{m} \sum_{c \in C_j} (2S_{ij}^c(\mathbf{u}_i - \mathbf{u}_c) + \sqrt{B_1^i}\mathbf{o}_{ij}^c) + 2\lambda\mathbf{u}_i. \quad (10)$$

Theorem 3.3. *If each element in $\mathbf{o}_{f,i}$ and $\mathbf{o}_{c,i}$ is independently and randomly selected from $L(4\sqrt{d}/\varepsilon)$, the derived U will satisfy $\epsilon-$differential privacy.*

Proof. According to Propositions 2.5 and 2.6, both of our noise generation methods can get the result that $\sum_{j=1}^{m}\sum_{f\in F_i^+}\mathbf{o}_{ij}^f = \mathbf{o}_{f,i} \sim L(4\sqrt{d}/\varepsilon)$ and $\sum_{j=1}^{m}\sum_{c\in C_j}\mathbf{o}_{ij}^c = \mathbf{o}_{c,i} \sim L(4\sqrt{d}/\varepsilon)$. During the process of updating the target user's latent vector $\mathbf{u}_i$, the summation of all the random noise from all the other users can be represented as follows:

$$\mathbf{o}_i = \sqrt{B_1^i}\mathbf{o}_{f,i} + \sqrt{B_1^i}\mathbf{o}_{c,i} \sim \sqrt{B_1}\left(L\left(\frac{4\sqrt{d}}{\varepsilon}\right) + L\left(\frac{4\sqrt{d}}{\varepsilon}\right)\right).$$

Since the infinite divisibility of classical Laplace distribution is stable, which is described in Proposition 2.5, we can get that each element in $\mathbf{o}_i = \{\mathbf{o}_{i1}, \mathbf{o}_{i2}, \ldots, \mathbf{o}_{id}\}$ is distributed as $L(2\sqrt{d}/\varepsilon)$.

Let D, D' be two neighboring datasets that only differ by one record $\mathbf{u}_f$ (or $\mathbf{u}_c$) and $\tilde{\mathbf{u}_f}$. Since from the different inputs, we obtain the same output U that is obtained after convergence, we have that $\frac{\partial J(D)}{\partial \mathbf{u}_i} = \frac{\partial J(D')}{\partial \mathbf{u}_i} = 0$, where J is as Equation (10) with an additional term $2\sum_{j=1}^{m}\mathbf{I}_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{ij})\mathbf{v}_j$, which is from the matrix factorization part. Thus, we have that

$$\mathbf{o}_i + 2\sum_{f\in F_i} S_{ij}^f(\mathbf{u}_i - \mathbf{u}_f) = \tilde{\mathbf{o}_i} + 2\sum_{f\in F_i} \tilde{S_{ij}}^f(\mathbf{u}_i - \tilde{\mathbf{u}_f}).$$

Since only the records $\mathbf{u}_f$ and $\tilde{\mathbf{u}_f}$ make a difference, we get that

$$\mathbf{o}_i - \tilde{\mathbf{o}_i} = 2\left(S_{ij}^f - \tilde{S_{ij}^f}\right)\mathbf{u}_i + 2S_{ij}^f\mathbf{u}_f - 2\tilde{S_{ij}^f}\tilde{\mathbf{u}_f}.$$

Considering that $S_{ij}^f \in [0,1]$, $||\mathbf{u}_i|| \le 1$, we have that

$$||\mathbf{o}_i - \tilde{\mathbf{o}_i}|| \le 4.$$

For each vector $\mathbf{u}_i$ of the derived $\mathbf{U}$, we can get that

$$\frac{Pr[\mathbf{u}_i|D]}{Pr[\mathbf{u}_i|D']} = \frac{\prod_{l=1}^{d} Pr(o_{il})}{\prod_{l=1}^{d} Pr(\tilde{o_{il}})} = e^{-\frac{\varepsilon\sum_{l=1}^{d}|o_{il}|}{4\sqrt{d}}} / e^{-\frac{\varepsilon\sum_{l=1}^{d}|\tilde{o_{il}}|}{4\sqrt{d}}} = e^{\frac{\varepsilon\sum_{l=1}^{d}(|o_{il}|-|\tilde{o_{il}}|)}{4\sqrt{d}}} \le e^{\frac{\varepsilon\sqrt{d\sum_{l=1}^{d}(|o_{il}|-|\tilde{o_{il}}|)^2}}{4\sqrt{d}}} = e^{\frac{\varepsilon\sqrt{d}||o_{il}-\tilde{o_{il}}||}{4\sqrt{d}}} \le e^{\varepsilon}.$$

Thus, we can provide the privacy guarantee consistent with $\varepsilon-$differential privacy for the derived $\mathbf{U}$. □

Following the above privacy-preserving recommendation design based on the Laplace Difference Privacy Mechanism, ID-SR can handle all the threats presented in Section 3.2 and provide better recommendation results while ensuring that the users' sensitive data, e.g., the actual rating values and latent vectors, are not leaked, protecting the user's privacy.

3.6 Our Scheme ID-SR

Combining the social recommendation method and perturbation scheme we proposed above, ultimately, ID-SR under classification optimization scheme (denoted as ID-SR(k), k indicates the number of categories) aims to solve the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V}} J = \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{I}_{ij}^1\left(\left(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j\right)^2 + \mathbf{v}_j^T\mathbf{o}_{1,j}^i\right) + \sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{I}_{ij}^2\left(\left(\mathbf{R}_{ij} - \mathbf{u}_i^T\mathbf{v}_j\right)^2 + \mathbf{v}_j^T\mathbf{o}_{2,j}^i\right) + \cdots$$

$$+\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{I}_{ij}^{K}\left(\left(\mathbf{R}_{ij}-\mathbf{u}_i^T\mathbf{v}_j\right)^2+\mathbf{v}_j^T\mathbf{o}_{K,j}^{i}\right)+\alpha\sum_{i=1}^{n}\sum_{j=1}^{|V_i|}\sum_{f\in F_i}\left(S_{ij}^{f}||\mathbf{u}_i-\mathbf{u}_f||_F^2+\mathbf{u}_i^T\sqrt{B_1^i}\mathbf{o}_{ij}^{f}\right)$$
$$+\alpha\sum_{i=1}^{n}\sum_{j=1}^{|V_i|}\sum_{c\in C_j}\left(S_{ij}^{c}||\mathbf{u}_i-\mathbf{u}_c||_F^2+\mathbf{u}_i^T\sqrt{B_1^i}\mathbf{o}_{ij}^{c}\right)+\lambda\left(||\mathbf{U}||_F^2+||\mathbf{V}||_F^2\right), \tag{11}$$

where α is a scalar to control the effect of social regularization. When there is no categorization of the ratings, i.e., the number of categories is 1, the above formula can be used as the objective function of the undifferentiated scheme. Thus, we will not repeat it. In fact, the undifferentiated perturbation scheme (denoted as ID-SR(non)) can be viewed as a special form of classification optimization scheme. Therefore, for the sake of convenience, the subsequent explanation focuses solely on the classification optimization scheme. In ID-SR, we use the gradient descent algorithm to iteratively optimize the user latent vector $\mathbf{u}_i$ and the item latent vector $\mathbf{v}_j$, obtaining the user profile matrix $\mathbf{U}$ and item profile matrix $\mathbf{V}$. The gradients of Equation (11) with respect to $\mathbf{u}_i$ and $\mathbf{v}_j$ are given as follows:

$$\frac{\partial J}{\partial \mathbf{v}_j}=\sum_{k=1}^{K}\sum_{i=1}^{n}\mathbf{I}_{ij}^{k}\left(2\left(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij}\right)\mathbf{u}_i+\mathbf{o}_{k,j}^{i}\right)+2\lambda\mathbf{v}_j \tag{12}$$

$$\begin{aligned}\frac{\partial J}{\partial \mathbf{u}_i}&=2\sum_{j=1}^{m}\mathbf{I}_{ij}\left(\mathbf{u}_i^T\mathbf{v}_j-\mathbf{R}_{ij}\right)\mathbf{v}_j+2\lambda\mathbf{u}_i\\&\quad+\alpha\sum_{j=1}^{m}\sum_{f\in F_i}\left(2S_{ij}^{f}(\mathbf{u}_i-\mathbf{u}_f)+\sqrt{B_1^i}\mathbf{o}_{ij}^{f}\right)+\alpha\sum_{j=1}^{m}\sum_{c\in C_j}\left(2S_{ij}^{c}(\mathbf{u}_i-\mathbf{u}_c)+\sqrt{B_1^i}\mathbf{o}_{ij}^{c}\right).\end{aligned} \tag{13}$$

To protect the users' information against the untrusted recommendation server, users, and attackers, our improved privacy-preserving social recommendation scheme, ID-SR, will follow the process shown in Figure 2

In our system, the recommendation server holds the item sets. Each user holds its own ratings and social relationships. The overall process is divided into three phases: initialization, optimization, and result generation. In the initialization phase, the recommendation server initializes the item profile matrix $\mathbf{V}$ and sends it to all users, whereas the user initializes its own latent vector $\mathbf{u}_i$. After this, users will interact with each other and the server to work together to complete the optimization phase. Specifically, in each iteration, the recommendation server aggregates perturbation data from each user according to Equation (12) in order to update the item profile matrix and send it to each user. Users, in turn, aggregate perturbation data from other users according to Equation (13) to update their own latent vector until convergence. Finally, using the converged $\mathbf{V}$ and $\mathbf{u}_i$, the user computes the final result $\hat{\mathbf{R}}_i=\mathbf{u}_i^T\mathbf{V}$ and obtains the recommendation result. The algorithm of ID-SR is described in Algorithm 1.

ID-SR is trained using rating data and social relationship data. A suitable differential privacy scheme is designed to protect users' data. At the end of the algorithm, a matrix of predicted ratings for each user is returned; based on this, items will be selected to recommend for the user.

Theorem 3.4. *ID-SR satisfies ε−differential privacy.*

Proof. In our scheme ID-SR, we optimize the item profile matrix $\mathbf{V}$ and user profile matrix $\mathbf{U}$ based on Equation (12) and Equation (13), respectively, instead of taking joint optimization. Based

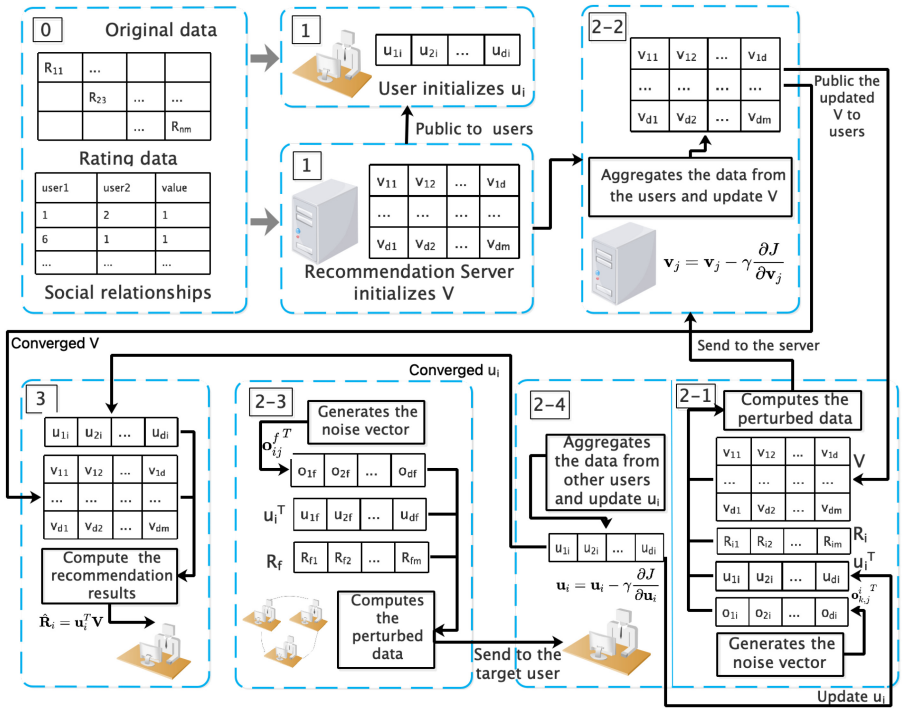


Fig. 2. The working process of ID-SR.

on Theorems 3.1, 3.2, and 3.3, both derived V and U under different versions of ID-SR satisfy ε-differential privacy. Thus, our scheme ID-SR satisfies ε-differential privacy and can resist attacks such as stealing and inferencing on user ratings and latent vectors by different attackers in the overall process. □

4 EXPERIMENTAL EVALUATION

In this section, our primary objective is to evaluate the effectiveness of our proposed scheme by answering the following two questions:

(1) Can our proposed improved social recommendation method I-SR yield better recommendation results?
(2) Can ID-SR achieve a more optimal balance between privacy protection and recommendation effectiveness?

To investigate these two questions, we implement I-SR, three forms of ID-SR, and perform the subsequent experimental setups.

4.1 Experimental Settings

In this article, we primarily focus on two publicly available social recommendation databases: CiaoDVD[2] and Epinions.[3] Both databases contain two files: the rating file, which includes users'

[2] http://www.ciao.co.uk/
[3] http://www.epinions.com/

ALGORITHM 1: Improved Privacy-Preserving Social Recommendation Based on Differential Privacy (ID-SR)

Input: $J, \varepsilon, \gamma, \alpha, \lambda$, User ratings $\mathbf{R}$ and social relationships data
Output: $\hat{\mathbf{R}}$

Users initialize their own latent vector $\mathbf{u}_i$, Recommendation Server initializes item profile matrix $\mathbf{V}$
while not converge **do**
  //Recommendation server updates $\mathbf{v}_j$
  for $j = 1, \ldots, m$ **do**
    for $k = 1, \ldots, K$ **do**
      for i in $R_{k,j}$ **do**
        Send $2(\mathbf{u}_i^T \mathbf{v}_j - \mathbf{R}_{ij})\mathbf{u}_i + \mathbf{o}_{k,j}^i$ to the recommender
      end for
    end for
    Update $\mathbf{v}_j = \mathbf{v}_j - \gamma \frac{\partial J}{\partial \mathbf{v}_j}$
  end for
  //User updates $\mathbf{u}_i$
  for $i = 1, \ldots, n$ **do**
    for $j = 1, \ldots, m$ **do**
      User i send $\sqrt{B_1^i}$ to the users in $F_i \cup C_j$
      for f in F_i **do**
        Send $\sqrt{B_1^i}\mathbf{o}_{ij}^f - 2S_{ij}^f \mathbf{u}_f$ to the user i
      end for
      for c in C_j **do**
        Send $\sqrt{B_1^i}\mathbf{o}_{ij}^c - 2S_{ij}^c \mathbf{u}_c$ to the user i
      end for
    end for
    Update $\mathbf{u}_i = \mathbf{u}_i - \gamma \frac{\partial J}{\partial \mathbf{u}_i}$
  end for
end while

return $\hat{\mathbf{R}} = \mathbf{UV}$

ratings of items on a scale from 1 to 5, and the social relationship file, which captures the relationships between users, represented by a directed binary value, i.e., a social value of 1 indicates that user i has a relationship with user f, whereas 0 signifies no relationship. Statistical information for the datasets is provided in Table 2.

In our experiments, we use five-fold cross-validation to partition the dataset and evaluate the scheme. **Mean Absolute Error (MAE)** and **Root Mean Square Error(RMSE)** are used as the metrics, which are defined as follows:

$$MAE = \frac{\sum_{(i,j)\in R_{test}} |\hat{\mathbf{R}_{ij}} - \mathbf{R}_{ij}|}{|R_{test}|}$$

$$RMSE = \sqrt{\frac{\sum_{(i,j)\in R_{test}} (\hat{\mathbf{R}_{ij}} - \mathbf{R}_{ij})^2}{|R_{test}|}},$$

where R_{test} is the set of ratings in the testing set, and smaller MAE and RMSE usually indicate better performance for recommendation effectiveness.

Table 2. Statistics of Datasets

Dataset	Users	Items	Ratings	Relationships
CiaoDVD	17,615	16,121	72,665	40,133
Epinions	49,290	139,738	664,824	487,181

In terms of parameters, some of them are set uniformly in the following experiments. We set the number of latent factors $d = 10$, matrix regularization term coefficient $\lambda = 10^{-3}$, and social regularization term coefficient $\alpha = 10^{-2}$. Except for the experiment that explores the effect of learning rate, we set $\gamma = 10^{-3}$. More descriptions of the parameters, especially about the effect of certain parameters, will be given in the following experiments.

4.2 Effectiveness of the New Method

To answer the first question, which pertains to evaluating the efficacy of I-SR, we conduct a comparative experiment between I-SR and the following methods.

- **Funk_svd** [9]**:** Basic Matrix Factorization Model. It decomposes the matrix **R** into two low-dimensional matrices and uses them to predict the user ratings of items. A regularity is added to the objective function to control the model variance.
- **Social_reg** [19]**:** Recommendation Systems with Social Regularization. It introduces a social regularization term based on matrix factorization. Complementing social relationships can be an effective way to improve recommendations.

The results at $\gamma = 10^{-2}$ and 10^{-3} are shown in Table 3. As can be seen from the table, I-SR outperforms the other two algorithms at different learning rates. Since all of the above algorithms use the same matrix factorization model, it can be concluded that our proposed improved social regularization term can effectively improve the recommendation results. In fact, I-SR can be considered as the version of ID-SR without differential privacy. From the above experiment results, it can also be seen that the learning rate does not significantly affect the algorithms without privacy preserving. In other words, smaller learning rates do not significantly improve the recommendation effectiveness of the algorithms without differential privacy.

4.3 Evaluation of ID-SR

In order to answer the second question and evaluate the effectiveness of ID-SR, we implement different versions (ID-SR(non), ID-SR(2), and ID-SR(3)) of our scheme and select some recent research (DPMF [13], PrivSR [21]) for comparison. The brief introductions on these schemes are as follows.

- **DPMF** [13]**:** Differentially Private Matrix Factorization. It is based on the basic factorization model and uses objective perturbation to ensure that the final item profiles satisfy differential privacy. It solves the challenge of decomposing the noise component into small pieces.
- **DPMF(modify):** A modified version of DPMF that we implemented. While the original DPMF uses an undifferentiated noise generation scheme that provides the same perturbation to all ratings, DPMF(modify) uses a classification optimization scheme under two categories, classifying ratings into private and public, and giving different perturbations to them.
- **PrivSR** [21]**:** Personalized Privacy-Preserving Social Recommendation. It is based on the social recommendation model and uses objective perturbation, allocating different noise magnitudes to personalized sensitive and non-sensitive ratings.

Table 3. Comparisons for Illustrating the Effectiveness of I-SR

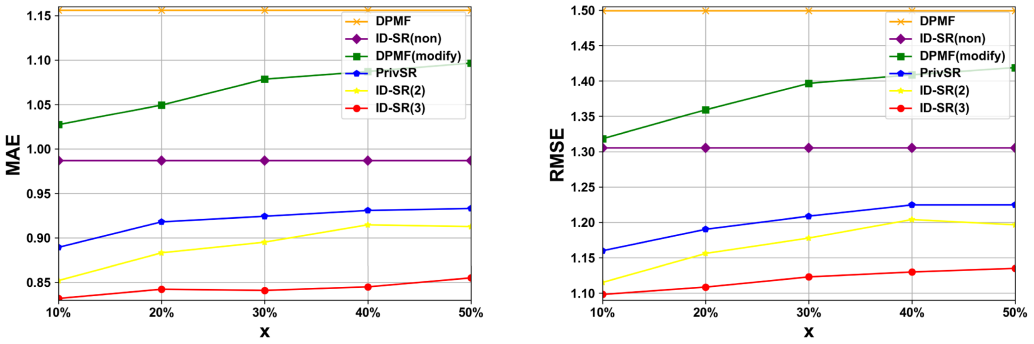
Dataset	Learning Rate	Method	MAE	RMSE
CiaoDVD	10^{-2}	Funk_svd	0.78176	1.02688
		Social_reg	0.76845	1.1616
		I-SR	0.7401	0.98233
	10^{-3}	Funk_svd	0.77806	1.02627
		Social_reg	0.76636	1.01491
		I-SR	0.74634	0.9861
Epinions	10^{-2}	Funk_svd	0.83209	1.09294
		Social_reg	0.82053	1.07165
		I-SR	0.81188	1.0646
	10^{-3}	Funk_svd	0.832	1.0924
		Social_reg	0.81925	1.05876
		I-SR	0.81088	1.05472

—**ID-SR(non):** Version of our ID-SR under an Undifferentiated Noise Generation scheme. Based on our proposed improved social recommendation algorithm I-SR, the objective function is perturbed based on the Laplace differential privacy mechanism, and all ratings are equally protected.
—**ID-SR(2):** Our scheme ID-SR under the Classification Optimization scheme with two categories. Since PrivSR is a 2-category algorithm, we also implemented a 2-classification version of DPMF and ID-SR for a fair comparison to validate the effectiveness of our scheme.
—**ID-SR(3):** Our scheme ID-SR under the Classification Optimization scheme with three categories.

Among these schemes based on differential privacy, given the presence of multiple parameters that could influence the outcomes, several experiments were designed through the controlled variable method to make a comprehensive evaluation of our scheme.

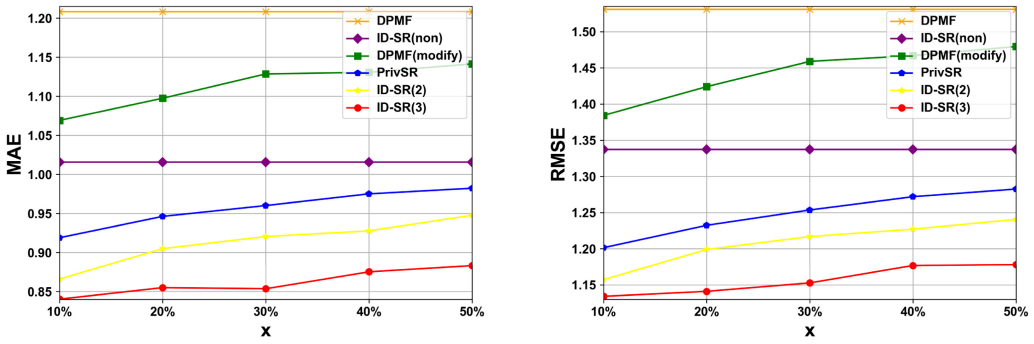
First, we would like to start by taking a general look at the performance of ID-SR and assessing the impact of categorization on the results. Under both datasets, we fix the learning rate $\gamma = 10^{-3}$, privacy budget $\varepsilon = 1$, and vary the percentage of the first and second categories. Among these schemes, DPMF and ID-SR(non) are both undifferentiated noise generation approaches that remain unaffected by data classification. At the same time, DPMF(modify), PrivSR, and ID-SR(2) are both two-classification schemes wherein the historical rating is categorized into private and public groups. To simulate the user's actions, we randomly select x percent of the ratings as private and the remaining 100-x as public. ID-SR(3) is the three-classification implementation of our scheme. x corresponds to the proportion of private and partially public data, whereas the fully public data percentages are given by 100-2x. We vary x as $\{10, 20, \ldots, 50\}$ and the results are shown in Figure 3 and Figure 4.

It can be seen that under different datasets, compared with DPMF, our undifferentiated noise generation scheme ID-SR(non) can lead to a significant improvement in recommendation effectiveness, and the recommendation accuracy even exceeds that of DPMF(modify), which has been improved by classification optimization. At the same time, compared with other schemes with the same two-classification optimization, our scheme ID-SR(2) is outperforming. Our three-classification scheme, ID-SR(3), can provide the best recommendation results among all of the methods above while providing privacy protection for the data through more fine-grained perturbation imposition, which can illustrate the effectiveness of our improved social regularization term and the different perturbation schemes proposed. In this experiment, an increase in the proportion



(a) MAEs of different schemes (b) RMSEs of different schemes

Fig. 3. Comparison on CiaoDVD under the same privacy budget and different data divisions.



(a) MAEs of different schemes (b) RMSEs of different schemes

Fig. 4. Comparison on Epinions under the same privacy budget and different data divisions.

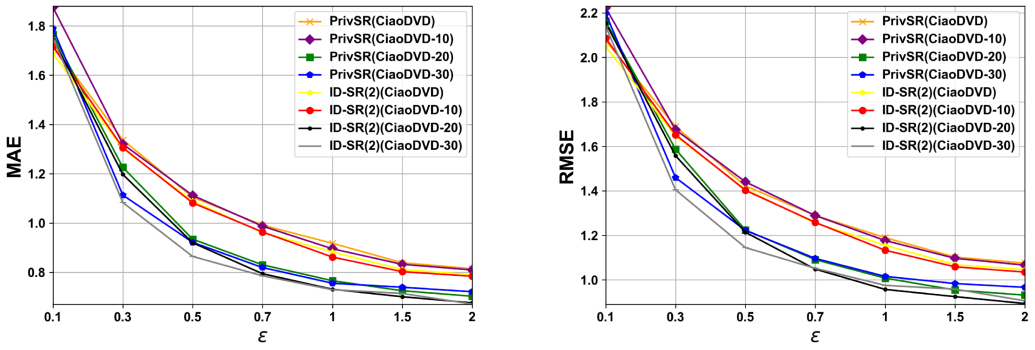
Table 4. Recommendation Results at Different Learning Rates

Method	Learning Rate	MAE	RMSE
ID-SR(2)	10^{-3}	1.68593	2.04711
	10^{-4}	1.30603	1.65864
	10^{-5}	0.98246	1.30258
PrivSR	10^{-3}	1.71936	2.07609
	10^{-4}	1.3063	1.65712
	10^{-5}	1.00232	1.32527

of private and partially public data, i.e., applying a larger proportion of loud noise to the dataset as a whole, brings about a small but insignificant change in accuracy. Later in this article, we will set x = 20%, i.e., select 20% of the data as private (20% each of private and partially public data in ID-SR(3)), as a representative scenario to evaluate the performance of ID-SR.

During our experiments, we found that the recommendation effectiveness of schemes with privacy preservation is strongly influenced by certain parameters. Thus, in a second step, we want to explore the impact of different factors on our schemes. First, we explore the impact of the learning rate. The results of PrivSR and ID-SR(2) under the CiaoDVD dataset with privacy budget $\varepsilon = 0.1$ at different learning rates are shown in Table 4.

As we have seen, smaller learning rates can lead to significant improvements in recommendation results but will come at the cost of longer training times. For a practical recommender, fast



(a) MAEs on different datasets (b) RMSEs on different datasets

Fig. 5. Recommendation effectiveness under different datasets and privacy budget.

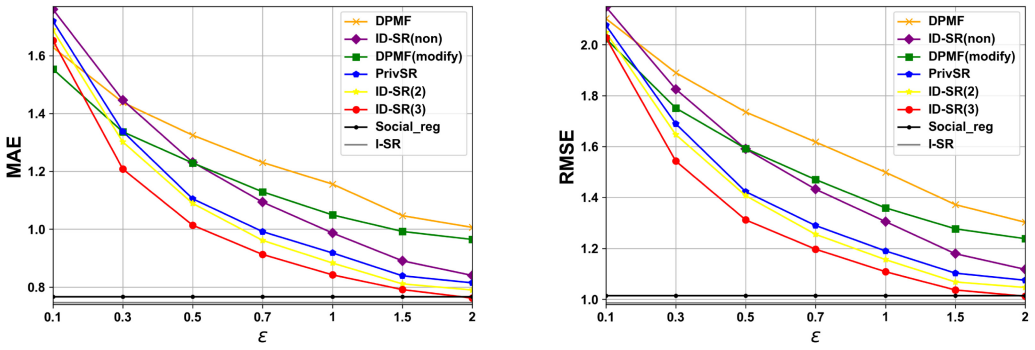
recommendations are essential. Thus, pursuing better results by tuning the learning rate is not what we are after. Our focus is on longitudinal comparisons with some parameters to illustrate the performance of our algorithms, not on how to tweak the parameters to make recommenders perform better. In subsequent experiments, we will fix the learning rate to 10^{-3} to evaluate our scheme.

We also found that the effect of schemes with differential privacy based on objective perturbation is strongly influenced by the number of ratings of each item in the dataset. Thus, we conducted the following experiments. We filtered the CiaoDVD dataset by sequentially filtering items rated less than 10, 20, and 30 and forming three datasets based on this, CiaoDVD-10, CiaoDVD-20, and CiaoDVD-30. Under different datasets, we set $\gamma = 10^{-3}$ and $x = 20\%$ to evaluate the performance of PrivSR and ID-SR(2) under different privacy budgets. The experimental results are shown in Figure 5.

Overall, the dataset with the original unfiltered items performs better when the privacy budget is very small ($\varepsilon = 0.1$). In other cases, the dataset that filters more items shows significant improvement in recommendation results. This is in line with our predictions because, first of all, items with more ratings bring better forecasts due to the fact that they can be predicted by a larger number of users who can predict their features and, thus, describe their hidden vectors more easily. Second, for items with only very few ratings, the noise perturbation added to them is huge. Indeed, for an item, when the privacy budget is small, e.g., $\varepsilon = 0.3$, the noise imposed on it will approximately follow $L(60)$. When the rating number of an item is small, e.g., only 5, the perturbation applied to the user's hidden vector approximately conforms to $L(12)$. This noise is relatively large, especially compared with the user's latent vector $\mathbf{u}_i$, as $||\mathbf{u}_i|| \leq 1$. Excessive noise significantly impacts the recommendation, and there is no need to add them to the user's hidden vector. Fortunately, modern recommendation systems typically have a substantial user base that rates each item much more frequently than our test dataset, resulting in the method performing well on large-scale datasets. Additionally, when a few items have a limited number of ratings, indicating the occurrence of the item cold-start problem, the recommendation results can be further supplemented based on the content of the item descriptions.

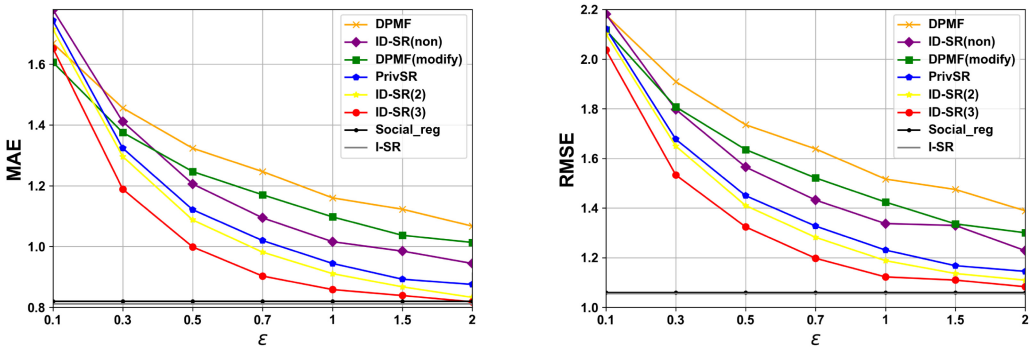
Based on the above discussion, we finally evaluate the impact of different privacy budgets on the effectiveness of each scheme under the same learning rate and dataset. We set ε from 0.1 to 2; the results are shown in Figure 6 and Figure 7.

At $\varepsilon = 0.1$, DPMF has the best recommendation accuracy because other schemes with social recommendation methods require additional protection of the user's hidden vectors, and more



(a) MAEs of different schemes (b) RMSEs of different schemes

Fig. 6. Comparison on CiaoDVD under different privacy budgets.



(a) MAEs of different schemes (b) RMSEs of different schemes

Fig. 7. Comparison on Epinions under different privacy budgets.

noise needs to be added. Other than that, our scheme performs better under other privacy budgets, and the finer categorization scheme can significantly improve the recommendation results. The recommendation accuracy of ID-SR(3) is even close to Social_reg when the privacy budget $\varepsilon = 2$. This demonstrates that our scheme performs better and strikes a better balance between privacy protection and recommendation effectiveness.

5 CONCLUSIONS AND FUTURE WORK

In this article, we propose a novel privacy-preserving social recommendation scheme for trustworthy AI called ID-SR, which aims to provide better recommendations while protecting user data. It is a solution to the AI security problem in recommender systems. ID-SR synthesizes the impact of social relationships and users with similar preferences on the target user's preferences. It considers user similarity at a fine-grained level for each item. We enhance the traditional matrix decomposition algorithm for social regularization terms and introduce a new matrix decomposition-based social recommendation method: I-SR. To protect users' private data from potential theft or inference by untrustworthy servers, users, and external attackers, we develop a differential privacy-preserving scheme specifically adapted to I-SR. Our approach leverages the infinite divisibility and stability properties of the Laplace distribution. We introduce perturbations to the objective function, conforming to the Laplace mechanism, in order to protect the generated user latent vectors and item latent vectors. We implement various versions of ID-SR by dividing the historical rating

dataset in different ways. Through experimental evaluation, we demonstrate that ID-SR outperforms other approaches, delivering improved recommendation results while effectively preserving user data privacy.

In our future work, we plan to explore scenarios such as Top-N recommendation and address the challenge of novelty recommendation by considering the dynamic dataset problem. Additionally, we aim to investigate more robust privacy-preserving solutions, ultimately offering users more secure and personalized recommendation methods.

REFERENCES

[1] Zhiqi Bu, Jinshuo Dong, Qi Long, and Weijie J. Su. 2020. Deep learning with Gaussian differential privacy. *Harvard Data Science Review* 2020, 23 (2020), 10–1162.

[2] Joseph A. Calandrino, Ann Kilzer, Arvind Narayanan, Edward W. Felten, and Vitaly Shmatikov. 2011. "You might also like:" Privacy risks of collaborative filtering. In *2011 IEEE Symposium on Security and Privacy*. IEEE, 231–246.

[3] Fran Casino, Josep Domingo-Ferrer, Constantinos Patsakis, Domènec Puig, and Agusti Solanas. 2015. A k-anonymous approach to privacy preserving collaborative filtering. *J. Comput. System Sci.* 81, 6 (2015), 1000–1011.

[4] Kamalika Chaudhuri, Claire Monteleoni, and Anand D. Sarwate. 2011. Differentially private empirical risk minimization.*Journal of Machine Learning Research* 12, 3 (2011), 1069–1109.

[5] Hai Chen, Fulan Qian, Chang Liu, Yanping Zhang, Hang Su, and Shu Zhao. 2023. Training robust deep collaborative filtering models via adversarial noise propagation. *ACM Transactions on Information Systems* 42, 1 (2023), 1–27.

[6] Cynthia Dwork. 2006. Differential privacy. In *International Colloquium on Automata, Languages, and Programming*. Springer, 1–12.

[7] Cynthia Dwork, Aaron Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.

[8] Arik Friedman, Shlomo Berkovsky, and Mohamed Ali Kaafar. 2016. A differential privacy framework for matrix factorization recommender systems. *User Modeling and User-Adapted Interaction* 26 (2016), 425–458.

[9] Simon Funk. 2006. Funk_svd. (2006). http://sifter.org/simon/jour-nal/20061211.html

[10] Guibing Guo, Jie Zhang, Daniel Thalmann, and Neil Yorke-Smith. 2014. Etaf: An extended trust antecedents framework for trust prediction. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE, 540–547.

[11] Wenxing Hong, Hejia Zhang, and Jiacheng Zhu. 2022. FedHD: A privacy-preserving recommendation system with homomorphic encryption and differential privacy. In *International Conference on Computer Science and Education*. Springer, 581–594.

[12] Dongkun Hou, Jie Zhang, Jieming Ma, Xiaohui Zhu, and Ka Lok Man. 2021. Application of differential privacy for collaborative filtering based recommendation system: A survey. In *2021 12th International Symposium on Parallel Architectures, Algorithms and Programming (PAAP)*. IEEE, 97–101.

[13] Jingyu Hua, Chang Xia, and Sheng Zhong. 2015. Differentially private matrix factorization. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*. 1763–1770.

[14] Zach Jorgensen, Ting Yu, and Graham Cormode. 2015. Conservative or liberal? Personalized differential privacy. In *2015 IEEE 31st International Conference on Data Engineering*. IEEE, 1023–1034.

[15] Jinsu Kim, Dongyoung Koo, Yuna Kim, Hyunsoo Yoon, Junbum Shin, and Sungwook Kim. 2018. Efficient privacy-preserving matrix factorization for recommendation via fully homomorphic encryption. *ACM Transactions on Privacy and Security (TOPS)* 21, 4 (2018), 1–30.

[16] Hyeyoung Ko, Suyeon Lee, Yoonseo Park, and Anna Choi. 2022. A survey of recommendation systems: Recommendation models, techniques, and application fields. *Electronics* 11, 1 (2022), 141–188.

[17] Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgórski. 2001. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Number 183. Springer Science & Business Media, Berlin, Germany.

[18] Zhiwei Liu, Liangwei Yang, Ziwei Fan, Hao Peng, and Philip S. Yu. 2022. Federated social recommendation with graph neural network. *ACM Transactions on Intelligent Systems and Technology (TIST)* 13, 4 (2022), 1–24.

[19] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. 2011. Recommender systems with social regularization. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. 287–296.

[20] Rachana Mehta and Keyur Rana. 2017. A review on matrix factorization techniques in recommender systems. In *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)*. IEEE, 269–274.

[21] Xuying Meng, Suhang Wang, Kai Shu, Jundong Li, Bo Chen, Huan Liu, and Yujun Zhang. 2018. Personalized privacy-preserving social recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32. 3796–3803.

[22] Natalia Ponomareva, Hussein Hazimeh, Alex Kurakin, Zheng Xu, Carson Denison, H. Brendan McMahan, Sergei Vassilvitskii, Steve Chien, and Abhradeep Guha Thakurta. 2023. How to DP-fy ML: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research* 77 (2023), 1113–1201.

[23] Xun Ran, Yong Wang, Leo Yu Zhang, and Jun Ma. 2022. A differentially private matrix factorization based on vector perturbation for recommender system. *Neurocomputing* 483 (2022), 32–41.

[24] Hyejin Shin, Sungwook Kim, Junbum Shin, and Xiaokui Xiao. 2018. Privacy enhanced matrix factorization for recommendation with local differential privacy. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1770–1782.

[25] Jiliang Tang, Huiji Gao, Huan Liu, and Atish Das Sarma. 2012. eTrust: Understanding trust evolution in an online world. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 253–261.

[26] Yong Wang, Mingxing Gao, Xun Ran, Jun Ma, and Leo Yu Zhang. 2023. An improved matrix factorization with local differential privacy based on piecewise mechanism for recommendation systems. *Expert Systems with Applications* 216 (2023), 119457.

[27] Chuhan Wu, Fangzhao Wu, Yang Cao, Yongfeng Huang, and Xing Xie. 2021. FedGNN: Federated graph neural network for privacy-preserving recommendation. (2021). *arXiv:arXiv preprint arXiv:2102.04925*

[28] Yilin Xiao, Liang Xiao, Xiaozhen Lu, Hailu Zhang, Shui Yu, and H. Vincent Poor. 2020. Deep-reinforcement-learning-based user profile perturbation for privacy-aware recommendation. *IEEE Internet of Things Journal* 8, 6 (2020), 4560–4568.

[29] Guangquan Xu, Xinru Ding, Sihan Xu, Yan Jia, Shaoying Liu, Shicheng Feng, and Xi Zheng. 2023. Real-time diagnosis of configuration errors for software of AI server infrastructure. *IEEE Transactions on Dependable and Secure Computing* (2023).

[30] Guangquan Xu, Zhengbo Han, Lixiao Gong, Litao Jiao, Hongpeng Bai, Shaoying Liu, and Xi Zheng. 2022. ASQ-FastBM3D: An adaptive denoising framework for defending adversarial attacks in machine learning enabled systems. *IEEE Transactions on Reliability* 72, 1 (2022), 317–328.

[31] Guangquan Xu, Chen Qi, Wenyu Dong, Lixiao Gong, Shaoying Liu, Si Chen, Jian Liu, and Xi Zheng. 2022. A privacy-preserving medical data sharing scheme based on blockchain. *IEEE Journal of Biomedical and Health Informatics* 27, 2 (2022), 698–709.

[32] Jiaqi Zhai, Jian Liu, and Lusheng Chen. 2021. Extraction security of sequential aggregate signatures. *Chinese Journal of Electronics* 30, 5 (2021), 885–894.

[33] Shun Zhang, Laixiang Liu, Zhili Chen, and Hong Zhong. 2019. Probabilistic matrix factorization with personalized differential privacy. *Knowledge-Based Systems* 183 (2019), 104864.

[34] Shu Zhao, Ziwei Du, Jie Chen, Yanping Zhang, Jie Tang, and Philip S. Yu. 2023. Hierarchical representation learning for attributed networks. *IEEE Transactions on Knowledge and Data Engineering* 35, 3 (2023), 2641–2656.

[35] Shu Zhao, Wenyu Wang, Ziwei Du, Jie Chen, and Zhen Duan. 2023. A black-box adversarial attack method via Nesterov accelerated gradient and rewiring towards attacking graph neural networks. *IEEE Transactions on Big Data* 9, 6 (2023), 1586–1597.

[36] Hao Zhou, Geng Yang, Yang Xiang, Yunlu Bai, and Weiya Wang. 2021. A lightweight matrix factorization for recommendation with local differential privacy in big data. *IEEE Transactions on Big Data* 9, 1 (2021), 160–173.

[37] Tianqing Zhu, Gang Li, Yongli Ren, Wanlei Zhou, and Ping Xiong. 2013. Differential privacy for neighborhood-based collaborative filtering. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 752–759.

[38] Xue Zhu and Yuqing Sun. 2016. Differential privacy for collaborative filtering recommender algorithm. In *Proceedings of the 2016 ACM on International Workshop on Security and Privacy Analytics*. 9–16.

Received 15 September 2023; revised 15 September 2023; accepted 15 December 2023