Advanced Privacy-Preserving Federated Relationship Recommendation

Bin Xue
School of Information and Communications
National University of Defense Technology
Wuhan, China
xxbbxl@sina.com

Qinghua Zheng
Department of Computer Science and Technology
Xi'an Jiaotong University
Xi'an, China
qhzheng@xjtu.edu.cn

Weihu Zhao
School of Information and Communications
National University of Defense Technology
Wuhan, China
zhaoweihuandy@126.com

Zhinan Li
University Office
National University of Defense
Changsha, China
lizhinan@nudt.edu.cn

Tianrun Cui
Laboratory of Flexible Electronics Technology
Tsinghua University
Beijing, China
Ctrl19@mails.tsinghua.edu.cn

Xue Feng
Laboratory of Flexible Electronics Technology
Tsinghua University
Beijing, China
fenxue@tsinghua.edu.cn

***Abstract*—Federated relationship recommendation is a growing technology trend in future intelligent system. Over the recent years, federated recommendation has recorded an exponential growth, leading to millions of intelligent equipments, and still increasing. However, privacy-preserving is an intractable problem in modern digital systems. In this paper, a comprehensive review of privacy-preserving federated recommendation is presented. Particularly, three kinds of federated recommendation methods are reviewed based on matrix factorization, neural network and other algorithms. Moreover, five authoritative datasets and three general evaluation metrics are mainly described. Simultaneously, a serious of outstanding federated recommender systems with related research discussions are provided. Finally, some potential future works will be tried to improve the application performance.**

***Keywords*—Federated learning, feature recommendation, feature extraction, deep learning, privacy-preserving**

I. Introduction

With the expeditious growth of intelligent technologies, electronic information recommendation systems (RSs) are becoming increasingly popular in kinds of giant-platforms. Generally, to develop a practical RS system, it is necessary to collect massive pre-annotated samples in advance. However, because of the personal information preservation requirement, individual secret material cannot be shared haphazardly. Therefore, in many certain scenarios, it is not easy to collect sufficient pre-annotated data for RSs training using centralized training. With the information scattered in several clients, complex learning models can be trained efficiently by federated learning (FL). Particularly, the original clients are never leaved by the raw data which meets the privacy protection requirement. FL offers technical foundation to train RSs with the information scattered in several platforms. The federated learning process is illustrated in Fig. 1.

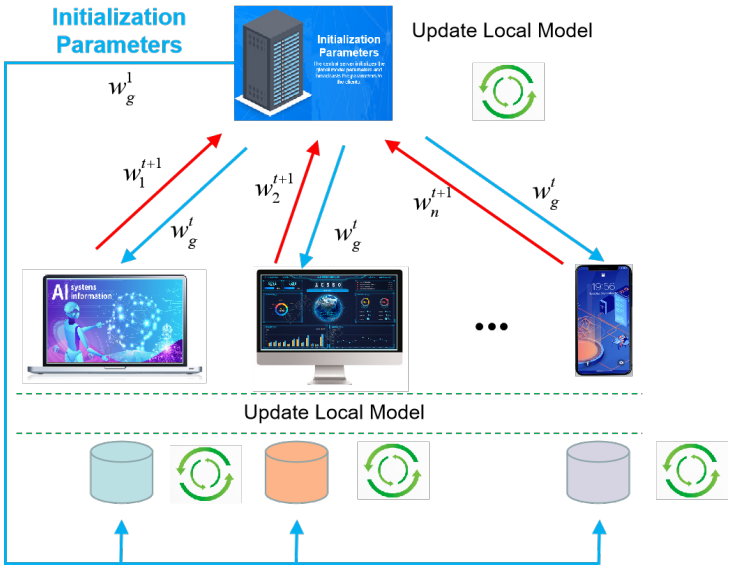

Fig. 1. The federated learning process.

2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) | 979-8-3503-2331-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAICA58456.2023.10405503

There have been proposed many Federated recommendation methods. In [1], a confederacy cooperative filtering is proposed using matrix factorization and confederacy random gradient descent. In [2], a model poisoning attack to federated recommender is developed to enhance objects' exposure ratio. Particularly, users' feature attributes are learned by public interactions, the poisoned gradients are generated accordingly by the attacker, and the malicious users can be controlled to transfer the poisoned gradients subtly. Recently, RSs based FL have attracted industries' and scholars' attention [3], which are utilized in many applications, including keyboard input, medical diagnosis, wireless communication, internet of vehicles, RS and so on.

In this paper, comprehensive review of privacy-preserving federated recommendation is proposed. The several main kinds of federated recommendation methods are described (in Section II). Five magisterial recommendation datasets and three common evaluation metrics are given, including recommendation accuracy, 10-fold Cross Validation, and Wilcoxon Rank-Sum Testing (in Section III). Finally, the conclusion and potential directions are offered.

II. Recent Development Progress

A. Federated Recommendation using Matrix Factorization

It is worth noting that consist with the classical FL system training, there are generally three processing stages in matrix factorization (MF) FedRec learning.

In [4], to improve the information extraction capability, a confederacy cooperative filtering (CCF) is proposed using MF and confederacy random gradient descent (FedSGD) polymerization. And a homomorphic encryption scheme is incorporated into CCF to additionally enhance the personal preservation performance. Extensive experiments are performed, showing the proposed methods achieves good performance.

Immediately after, a federated recommendater using recessive back-propagation MF and client property polymerization is developed by Lin et al. [5]. Customer preferences are strongly associated with product characteristics. A series of experiments verify that the presented model works better than FedRec using the classical matrix factorization.

B. Federated Recommendation using Neural Network

In these years, FedRec using neural network has been a research hot pot, and there are some related exquisite methods. In [6], Inspired by REPTILE meta learning, a simple yet effective federated recommendation technology is developed in a distributed manner. The model training and data transmission among the edge devices and parameter servers are performed in a special way in the cloud. Experiments show that the performance of both the data privacy-preserving and recommender personalization are enhanced obviously.

A multi-view recommendation system is constructed in [7] for generic content based on federated learning. It is worthy that the advantages of different models are integrated into one framework, and scattered data is effectively mined and utilized. Moreover, system accuracy is improved with multiple user-level featured data sources.

In [8], a federated recommendation model is developed with privacy-preserving graphical neural network (GNN). Particularly, a user-item graph enlargement scheme is presented to bring high-order user-item interactions. Some qualitative and quantative experiment results demonstrate that the method can be trained with decentralized user data collectively, and the high-order user-item interaction representations can be exploited simultaneously with privacy well protected. Fig. 2 describes different routes in graphical neural network.

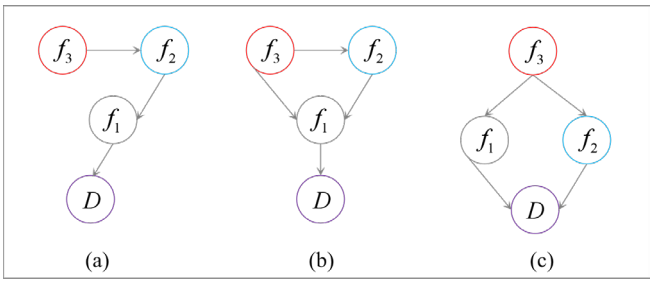

Fig. 2. Different routes in graphical neural network.

Non-independent distribution is a tricky issue in federated learning, a federated recommendation system is proposed in [9] using historical clustering to relieve this problem. Particularly, with a series of clever designs, the weighted average of different heterogeneous parameters has been effectively developed. Good performance is achieved by the proposed method.

C. Federated Recommendation using Other Algorithms

Suitable information can be offered by distributed recommender systems (DRSs) adaptively. However, it is difficult for DRSs to share information safely among agents. Therefore, a privacy-preserving DRS is developed in [10], and the service providers are regarded as distributed context learners. Moreover, a federated training strategy is adopted to learn efficient centralized system.

In [11], a federated point-of-Interest recommendation (PriRec) model is constructed. Particularly, local differential privacy strategy is presented to release dynamic information owing personal promises. PriRec is represented by factorization machine (FM). A safe gradient descent scheme is developed to learn a linear module and a feature interaction module to achieve model privacy.

In [12], a model poisoning attack to federated recommender is developed to enhance objects' exposure ratio. Particularly, users' feature attributes are learned by public interactions, the poisoned gradients are generated accordingly by the attacker, and the malicious users can be controlled to transfer the poisoned gradients subtly.

Usually, there are some adversarial attacks in federated learning, leading the shared information unreliable in multiple clients. Particularly, the Byzantine attacks are in the worst case. To solve this problem, in [13], a federated learning model is built using distributed low-rank matrix completion. Different Byzantine aggregation strategies are brought into the gradient descent scheme for performance improvement.

III. Datasets and Evaluation Metric

A. Authoritative Datasets

There are some authoritative datasets, such as

MovieLens100K [14], MovieLens1M [14], MovieLens20M [14], BookCrossing [15], and Frappe [16].

MovieLens datasets are movie recommendation datasets that include movie metadata information and user attribute information. According to the purpose of use, it can be divided into two categories: one type of dataset is suitable for advancing the latest research; A type of dataset is used for university research and educational research, which is divided into multiple datasets based on whether they have labels, time order, and dataset size. The main file details of these datasets are shown in Table I. Particularly, the links.csv document includes identifiers utilized to link to other movie data sources. Each line after the file header represents a movie, and the use of each movie resource is subject to the terms and conditions of each supplier. The movies.csv file contains movie information, with each line after the header representing a movie. The movie title can be manually entered or imported from it, with the release year written in parentheses. And all tags are included in the tags.csv file, and each line after the file header represents a tag applied by a user to a movie. The lines in this file are first sorted by userId, and then sorted by movieId in user. The meaning, value, and purpose of specific tags are determined by each user. Fig. 3 shows the movie search application example.

TABLE I MAIN DOCUMENTS DETAILS IN MOVIELENS DATASETS

Document	links.csv	movies.csv	ratings.csv	Readme.txt	tags.csv
Content	Different URLs for movie tags	Movie tags, titles, and movie types	The user's rating of the movie is an integer	Document Introduction	User reviews of movies

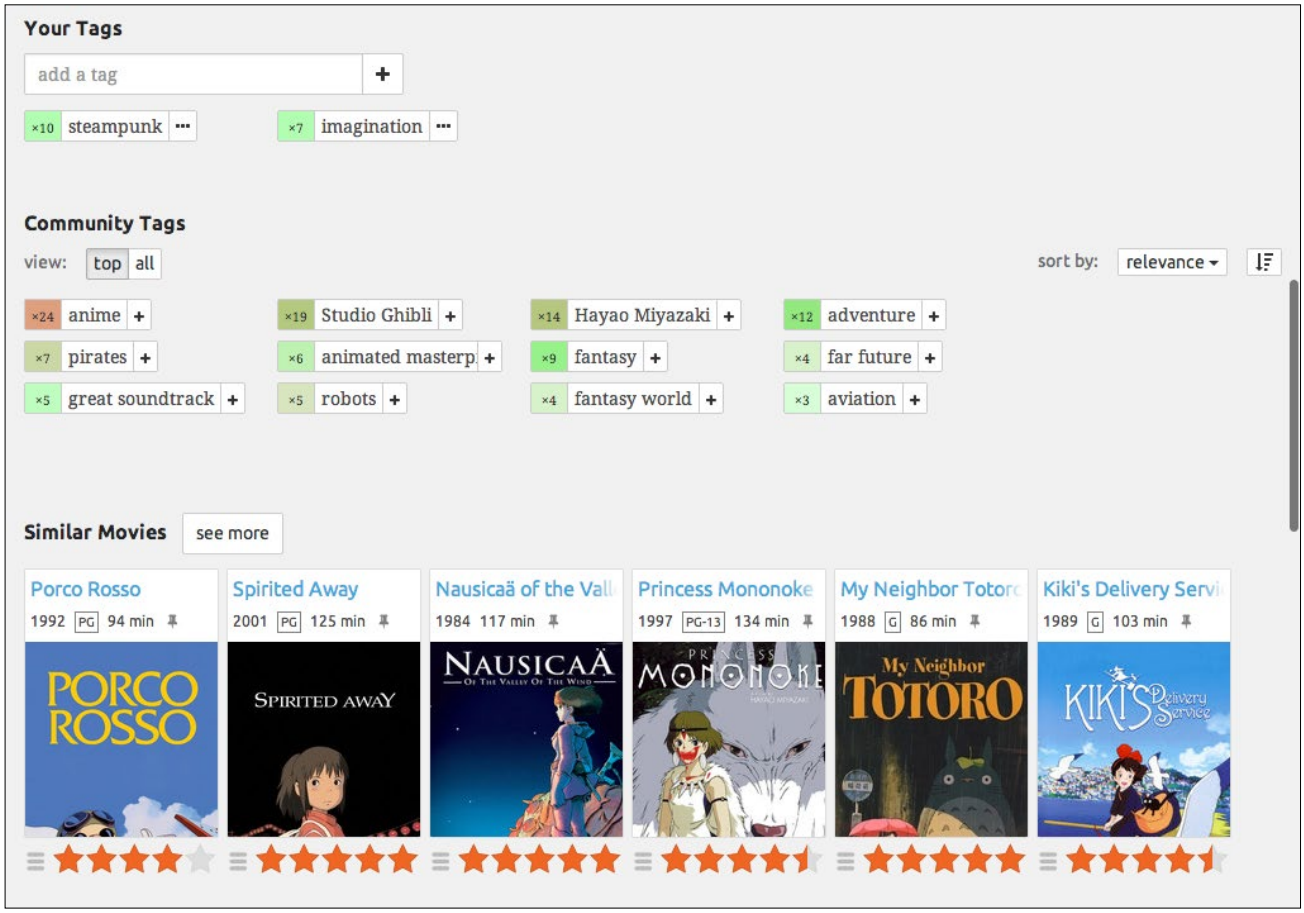

Fig. 3 Federated recommendation in movie search.

BookCrossing dataset is a popular internet book-grading community. There are 278858 members, 340000 books and 1157112 ratings. BookCrossing dataset consists of three categories: (1) BX-user, which contains user information, where the user ID has been anonymized and mapped to an integer. Except for the portion containing demographic data, all other fields contain NULL values. (2) BX-books, including the ISBN identification of the book, also provide content based information such as author, publication year, publisher, etc. In the case of multiple authors, only the first author is provided; And the dataset provides a URL to the cover image, with the relevant link directly pointing to Amazon. (3) BX-Bookrating, which includes book rating information, where ratings are divided into explicit values, values ranging from 1 to 10, and implicit values represented by 0. Furthermore, it is worthy that some kinds of customers' important personal information are revealed.

Frappe can be utilized for content recommendations, which includes 96203 app usage logs for users in diverse contents. Each log (customer number, algorithm number, content parameters) is transformed into a feature vector through a hot encoding. Fig. 4 illustrates the human resource management application example.

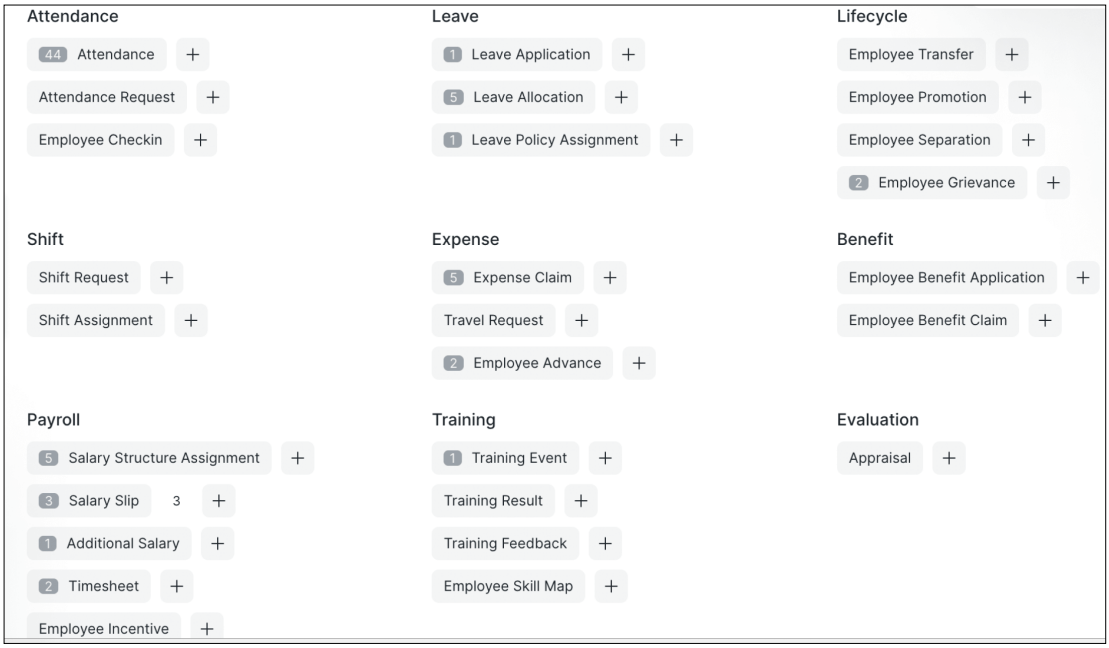

Fig. 4 Federated recommendation in human resource management.

B. Evaluation Metrics

1) Recommendation Precision

Recommendation precision can be applied to evaluate in DF_Rec, which is defined as

$$Recommendation\ Precision\ = TP/(TP\ +\ FP) \tag{1}$$

where TP and FP denote the true and false positive samples, respectively.

2) 10-fold Cross Validation (CV)

S_n is evenly and randomly classified into 10 folds of similar size $P=\{P_1,P_2,...,P_{10}\}$. $T_i=S_n/P_i$ denote P_i 's complementary dataset. The induction function $A(\cdot)$ contains classification effect related with T_i , $\psi_i=A(T_i)$ calculated the prediction error of P_i . Therefore, the 10-fold cross validation forecasting incorrect operator of $\psi=A(S_n)$ can be:

$$\hat{\varepsilon}_{10}(S_n,P)=\frac{1}{n}\sum_{i=1}^{10}\sum_{(x,c)\in P_i} l(c,\psi_i(x)) \tag{2}$$

where $l(i,j)=1$ if $i\neq j$ and zero otherwise.

3) Wilcoxon Rank-Sum Testing

Wilcoxon rank-sum testing is a non-parametric test that determines whether two independent samples were selected from items owing the same distribution. It can eveluate the independent but non-normally distributed data.

IV. Conclusions and Future Works

In this paper, comprehensive review of privacy-preserving federated recommendation is presented. The several main kinds of federated recommendation methods are described. Five magisterial recommendation datasets and three common evaluation metrics are given, including recommendation accuracy, 10-fold cross validation, and Wilcoxon rank-sum testing. Finally, the conclutson and potential future directions are offered.

In the future research, some works will be tried to improve the model performance: 1) The combination of shallow feature extraction and deep learning will be considered. 2) A small number of samples will be used to provide robust features and recommendation performance with less training time. 3) More ingenious combination/ fusion methods will be designed to connect deep learning frameworks and federated aggregation schemes more effectively, such as dynamic self-organizing module.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62106277, Scientific Research Plan of National University of Defense Technology (ZK21-39), Basic Strengthening Plan (2022-JCJQ-QT-049).

References

[1] X. Niu, X. Zhang, Z. Chu, and X. Li, "Federated Collaborative Filtering Recommendation Based on Semi-Homomorphic Encryption," 2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), in Chengdu, China, pp. 316-321, April 2023.

[2] J. Nie, Z. Zhao, L. Huang, W. Nie, and Z. Wei, "Cross-Domain Recommendation Via User-Clustering and Multidimensional Information Fusion," IEEE Trans. Multimedia, America, vol. 25, pp. 868-880, April 2023.

[3] X. Zhou, Z. Hu, and J. Huang, "Decentralized Gradient-Quantization Based Matrix Factorization for Fast Privacy-Preserving Point-of-Interest Recommendation," Journal of Artificial Intelligence Research, vol. 76, pp. 1019-1041, May 2023.

[4] M. Ammad-ud-din, E. Ivannikova, S. A. Khan, W. Oyomno, Q. Fu, K. Eeik Tan, and A. Flanagan, "Federated collaborative filtering for privacy-preserving personalized recommendation system," arXiv preprint arXiv:1901.09888, pp. 1-12, January 2019.

[5] G. Lin, F. Liang, W. Pan, and Z. Ming, "FedRec: Federated

Recommendation With Explicit Feedback," IEEE Intelligent Systems, America, vol. 36, pp. 21-30, October 2021.

[6] A. Jalalirad, M. Scavuzzo, C. Capota, and M. Sprague, "A Simple and Efficient Federated Recommender System," 6th Big Data Computing, Applications and Technologies (BDCAT), Auckland, New Zealand, pp. 53-58, December 2019.

[7] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, and Y. Liang, "Blockchain-Enabled Federated Learning Data Protection Aggregation Scheme With Differential Privacy and Homomorphic Encryption in IIoT," IEEE Trans. Ind Inform, America, vol. 18, pp. 4049-4058, June 2022.

[8] T. Sun, D. Li, and B. Wang, "Decentralized Federated Averaging," IEEE Trans. Pattern Anal, America, vol. 45, pp. 4289-4301, April 2023 .

[9] Z. Jie, S. Chen, J. Lai, M. Arif, and Z. He, "Personalized federated recommendation system with historical parameter clustering," Journal of Ambient Intelligence and Humanized Computing, pp. 1-11, 2022.

[10] P. Zhou, K. Wang, L. Guo, S. Gong, and B. Zheng, "A Privacy-Preserving Distributed Contextual Federated Online Learning Framework with Big Data Support in Social Recommender Systems," IEEE Transactions on Knowledge and Data Engineering, vol. 33, pp. 824-838, March 2021.

[11] C. Chen, J. Zhou, B. Wu, W. Fang, L. Wang, Y. Qi, and X. Zheng, "Practical Privacy Preserving POI Recommendation," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 11, pp. 1-20, 2020.

[12] D. Rong, S. Ye, R. Zhao, H. Ning Yuen, J. Chen, and Q. He, "FedRecAttack: Model Poisoning Attack to Federated Recommendation," 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, pp. 2643-2655, May 2022.

[13] X. He, Q. Ling, and T. Chen, "Byzantine-Robust Stochastic Gradient Descent for Distributed Low-Rank Matrix Completion," 2019 IEEE Data Science Workshop (DSW), Minneapolis, MN, USA, pp. 322-326, June 2019.

[14] F. Maxwell Harper, Joseph A. Konstan, "The movielens datasets: History and context," ACM Trans. Interact Intel, America, vol.5, pp. 1-19, February 2016.

[15] C. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen, "Improving recommendation lists through topic diversification," International Conference on World Wide Web. ACM, Chiba, Japan, pp. 22-32, 2005.

[16] L. Baltrunas, K. Church, A. Karatzoglou, and N. Oliver, "Frappe: Understanding the usage and perception of mobile app recommendations in-the-wild," arXiv preprint arXiv:1505.03014, pp. 1-10, May 2015.