IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) | 978-1-6654-0443-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/INFOCOMWKSHPS51825.2021.9484510

POI Recommendation with Federated Learning and Privacy Preserving in Cross Domain Recommendation

Li-e Wang[1], Yihui Wang[1], Yan Bai[2], Peng Liu[1*] and Xianxian Li[1*]
[1]Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin, China
[2]School of Engineering and Technology, University of Washington Tacoma, WA 98402, USA
wanglie@gxnu.edu.cn, wyh154950130@163.com, yanb@uw.edu, liupeng@gxnu.edu.cn, lixx@gxnu.edu.cn

***Abstract*—Point-of-Interest (POI) recommendation is one of the most popular recommendation methodologies. However, POI data is very sensitive and sparse. Users' reluctance to share their context information due to privacy concerns, along with the cold-start problem caused by data sparsity reduces recommendation efficiency. To address these issues, we propose a POI framework for cross-domain recommendation with federated learning and privacy protection features. It utilizes data in an auxiliary domain in users' interest analysis to alleviate the cold-start problem. Moreover, it applies federated learning by analyzing the users' historical data locally and encrypts latent feature distribution for knowledge migration to protect users' privacy. Experiments on real datasets have shown that our framework improves recommendation accuracy while preserving users' privacy as compared to convolutional neural network-based methods when analyzing users' comments.**

***Key words*—POI Recommendation, Cross-domain correlation, privacy-preserving, federated learning, machine learning**

I. INTRODUCTION

With the development of mobile localization technology, Point-of-Interest (POI) recommendations have attracted wide attention. POI recommendations can filter information based on users' interests to achieve effective recommendation and greatly improve users' experience while producing substantial economic benefits. Nowadays, many platforms support users releasing daily social network data while sharing their location "check-in" information and generating relevant comments. Unlike traditional product recommendations, people's check-in data is more sensitive and significantly sparser in POI recommendations. Check-in data not only reflects a person's location, but also reveals users' interests and other private details. For example, if a person frequently checks in to a hospital-related location over a period of time, we can infer that the user may have a health problem. In addition, POI recommendations have strict requirements on the time factor because people tend to checks in different POIs at different times during a day.

There are two main approaches to address these issues. One approach integrates POI check-in data with other information in POI field for auxiliary analysis and recommendations, such as user's comments [1]-[2] and social network data [3]; another approach applies cross-domain correlation to solve the cold-start problem, which is often based on the feature correlation between the cities where users often checks in [4]-[5]. These methods can alleviate the data sparsity and cold-start problem to some extent, but there still exist some limitations. Firstly, users' comments are sparser in the premise of POI check-in data is very sparse in itself. Users' comments can be affected by many other factors. They cannot fully indicate users' preference for a POI. Secondly, the city features-based methods may not apply to all cities. When two cities have fewer common features, the work of knowledge transfer becomes very difficult. Last but not least, very little relevant researches have considered the privacy issue of user check-in information. Especially when combining with other domains' data, such as in social networks, users' private information is more vulnerable to be revealed. Based on these analyses, we focus on addressing the privacy issues in cross-domain recommendation, which is more challenging than single-domain POI recommendations, and propose a POI recommendation with federated learning to solve the issue of data sparsity in cross domain scenarios. In particular, we design a recommendation framework with federated learning and privacy protection for cross-domain POI recommendations. Our federated learning mechanism is based on homomorphic encryption technology. The main contributions of this paper are as following.

1) To address the problem of data sparsity and cold-start users, we fuse the auxiliary domain data for machine learning to achieve transfer learning and improve the accuracy of recommendations. To the best of our knowledge, most of the existing works that carry out auxiliary data analysis are based on highly overlapping data for user features and cannot achieve personalized transfer learning. We consider partial overlaps of user features and use Multi-Layer Perceptron (MLP) to learn the latent features in different domains (i.e., the auxiliary domain and the target domain) to realize personalized knowledge transfer.

2) To address the privacy issues, we introduce federated learning and store user data locally for privacy protection. We construct a novel cross-domain POI recommendation framework based on federated learning and privacy protection (FL&PP-POI). To the best of our knowledge, this is the first research effort in introducing federated learning and considering data privacy in cross-domain POI recommendations.

3) We design a POI recommendation algorithm based on cross-domain correlation and privacy protection, which can implement effective recommendation while protecting users' privacy. Experimental results on Movielens and FourSquare datasets have shown that our method can effectively guarantee users' privacy while improving recommendation accuracy.

II. RELATED WORK

A. POI Recommendation

Common approaches to address data sparsity and the cold-start problem in POI recommendations consider time and

*corresponding author

978-1-6654-0443-3/21/$31.00 ©2021 IEEE

location factors. Zeng et al. [6] proposed a time-based collaborative filtering algorithm, which is based on the global similarity between users. Seo et al. [7] considered the correlation between the anchoring effect with POI recommendation, and assumed that the initial inputs do not exist. Other works have used users' comments in POI recommendations [8]-[10]. Buru et al. [8] also pointed out that traditional continuous POI models only focus on the physical distance between POIs and do not utilize the relationship between continuous POIs.

Currently, only very few works adopt cross-domain recommendation in the POI area. Most of them mainly focused on the correlation between social network data and POI check-in data. Interpersonal relationships can provide another analytical dimension for POI recommendations. Zhang et al. [11] proposed a probabilistic generation model to capture the social, sequential, temporal and spatial patterns of user check-in data. Other researchers considered cross-city recommendations. Li et al. [4] designed a transfer learning model based on the cities' common features, which can transfer users' real interest in a source city to a target city. Ding et al. [5] observed the relationship between tourists' check-in features in their hometown and tourism destination city, and the differences of check-in features between locals and tourists. They proposed a POI recommendation framework based on user check-in features transfer. Yet, other studies improve the POI recommendations' accuracy by combining data from another domain. Wen et al. [12] clustered users based on the word2Vec model and modeled the transfer behavior at the clustering level using additive Markov chain.

B. Privacy-preserving work in POI

Traditional privacy-preservation methods in POI recommendation include generalization [13]-[15], differential privacy [16] and encryption [17]. Wang et al. [13] protected users' privacy by learning user group preferences rather than individual preferences. Their recommendations are made by personalizing groups on the user's local device based on Nonnegative Matrix Factorization techniques. Li et al. [16] used Markov chains to model users' check-in sequence and protected their privacy by adding the weighted noise. Wang et al. [17] designed a PLSA encryption algorithm and proved that it is secure under a semi-honest protocol.

To the best of our knowledge, works in cross-domain POI recommendation which have considered privacy protection is very limited. Chen et al. [18] pointed that the main problem of the existing matrix factorization methods is centralized storage which will results in large computation and storage costs. At the same time, users' preferences can be leaked to attackers through centralized learning. A distributed training method was designed based on the random walk technology in which users' score information is protected locally to reduce computing and storage cost at servers. However, it does not involve the cross-domain data. Differing from existing studies, we consider the correlation between e-commerce and POI fields, and design an effective framework based on POI cross-domain recommendations with privacy protection.

III. PRELIMINARIES

A. Problem definition

When a new user checks into a platform, the platform does not have the user's historical data. The cold-start problem occurs. Supposing that there are n cold-start users in U, and the current POI of user is p, and the coordinates of p is (x, y), we can get the range $S = \pi \cdot r^2$, where r is the radius of the range S. There is a POIs set $P = \{p_1, p_2, \ldots, p_m\}$ within the range S. For any user $u_i, i \in \{1, \ldots, n\}$ in the set U, we need to generate a top-N recommended list L_i^m within range S while guaranteeing users' privacy. We call the range S as the "business circle".

B. LDA model

LDA (Latent Dirichlet Allocation Model) is a model of natural language processing, which is used to generate topic distribution for a document. It is a word bag model and does not consider the order of the words in the text.

Supposing that Z represents the topic assigned to the word in documents w, the probability that the LDA model generates a document W is:

$$p(W) = \sum_z p(z) \prod_{n=1}^{N} p(w_n \mid z) \qquad (1)$$

C. Homomorphic encryption

Homomorphic encryption is one of the most widely used privacy protection technologies at present. It can effectively protect data privacy and support computation of the encrypted data without decryption. In this work, we encrypt users' latent features using a homomorphic encryption algorithm to train a neural network. An improved activation function [19] and the encryption scheme in [20] are adopted to support both additive and multiplicative homomorphic for implementation effectiveness and efficiency.

IV. FL&PP-POI FRAMEWORK

In our proposed FL&PP-POI (Federated Learning and Privacy Preserving) framework, original data can be stored locally. After data analysis, the learned latent feature vectors are encrypted with homomorphic encryption technology and sent to a server. The server uses an improved activation function [19] to enable MLP method [21] to perform operations on ciphertexts and guarantee result validity.

A. The e-commerce field (Auxiliary domain A)

Every day users generate a large amount of behavioral data on websites such as clicks, browsing, purchases, and comments. The clicking and browsing behaviors are implicit feedback which are not easy to realize knowledge transfer. The comments may contain subjective information, which are often affected by other external factors. We, therefore, choose to analyze the word vector of users' purchase records.

Generally, cold-start users can be associated with other auxiliary domain to obtain relevant historical data. In this paper, the latent features of cold-start users and similar users in the auxiliary domain are analyzed by LDA model. By inputting latent feature distributions of similar users into a neural network,

we obtain a feature function which trains and then transfer the cold-start users' feature distributions in the auxiliary domain to the target domain. Since users always have different check-in features in different domains. What's more, we use cosine similarity to measure the similarity between different users. Finally, latent feature vectors are encrypted with homomorphic encryption technology and sent to the server. The detailed algorithm in the auxiliary domain A is shown in Algorithm 1.

Algorithm 1 Generate user's latent feature distribution

Input: The user's historical purchase word record $W_{u\in U}$; The similarity threshold *sim*

Output: The latent feature distribution of similar common users for a cold-start user in each quarter $\theta^{t_k}_{u\in SIM^{t_k}_{u_i}}$; The latent feature distribution of cold-start users in the auxiliary domain A $\theta^{A}_{u_i}$

1 Divides the user's historical data into four quarters $W^{t_k}_{u\in U}, k=\{1,2,3,4\}$

2 Divides 80% of user data into training sets and the remaining 20% into test sets

3 **For** each cold-start user u_i **do**

4 Calculates the similar common user set in four quarters $SIM^{t_k}_{u_i}, k=\{1,2,3,4\}$

5 Sample topics for each user from historical data in four quarters and get the distribution $\theta^{t_k}_{u\in SIM^{t_k}_{u_i}}$

6 **End For**

7 Sends the encrypted latent feature distribution to server

B. The POI field (Target domain B)

In the POI recommendation, we mainly analyze the common users' features in the target domain and the users' POI check-in features within the "business circle" as defined in Section III.A. Similar to the auxiliary domain, the common users' feature in a target domain is used to train the mapping function as the neural network's input. To increase accurate matches of POIs with interested users within the "business circle", we need to analyze the most similar POIs with the mapped cold-start users' distribution. The bigger the similarity between any POI's "business circle" feature distribution and the cold-start user is, the more the users' interests match. Keeping consistent with user data in e-commerce sites, the users' historical check-in data is also divided into four quarters. Since we recommend the POIs within the "business circle" around a POI that the user has not checked into, we only need to analyze the users' historical data within the "business circle". The detailed algorithm of the target domain B is shown in Algorithm 2.

Algorithm 2 Generate historical check-in distribution within the "business circle"

Input: The descriptive word for POI *Wp*; The latitude and longitude of the user's current POI (*x, y*); The threshold of the radius *r*

Output: The historical check-in feature within the "business circle" in four quarters of current POI $\theta^{t_k}_{p\in S}, k=\{1,2,3,4\}$

1 Calculates the geographic distance *d* between the current POIs and other POI within the "business circle"

2 **If** $d < r$:

3 Divides the historical check-in distribution within the "business circle" into four quarters $W^{t_k}_{p\in S}, k=\{1,2,3,4\}$

4 Sample topics for each user from historical data in four quarters within the "business circle" by LDA model $\theta^{t_k}_{p\in S}$

5 **End If**

C. The server

The server focuses on training the latent feature mapping function *F*. We chose the MLP method and use error back propagation to update the related gradient. We adopt an improved activation function [19] to train the encrypted latent feature vectors sent by both auxiliary domain *A* and target domain *B*. Finally, the server sends the cold-start users' mapped distribution to *B* to assist it in making a recommendation. The detailed algorithm is shown in Algorithm 3.

Algorithm 3 Generate mapped cold-start user's latent feature distribution

Input: The encrypted cold-start user's latent feature [$\theta^{A}_{u_i}$]; The encrypted latent feature of the cold-start user's similar common user [$\theta^{A}_{u\in SIM^{t_k}_{u_i}}$]in auxiliary domain A; The encrypted latent feature of the cold-start user's similar common user[$\theta^{B}_{u\in SIM^{t_k}_{u_i}}$] in target domain B

Output: The feature mapping function $F^{t_k}_{u}, k=\{1,2,3,4\}$; The mapped latent feature of the cold-start users $\widehat{\theta}^{t_k}_{u_i}$

1 **For** each cold-start user u_i **do**

2 Structure training set $T^{t_k}_{u_i}=\{[\theta^{A}_{u\in SIM^{t_k}_{u_i}}],[\theta^{B}_{u\in SIM^{t_k}_{u_i}}]\}$

3 Initialize the weights in MLP

4 **While** the loss function does not converge **do**

5 **For** each in the similar set **do**

6 Calculates the loss function value of MLP on the user u_i

7 Update the related weights and gradients

8 **End For**

9 **End While**

10 Calculates the mapped cold-start user's latent feature distribution $\widehat{\theta}^{t_k}_{u_i}$

11 **End For**

D. Generating a recommendation list

The target domain B obtain the cold-start users' mapped distribution $\widehat{\theta}^{t_k}_{T,u_i}$, and the check-in distribution $\theta^{t_k}_{T,u_i}$ analyzed in the target domain. We also use cosine similarity to measure the similarity $sim^{t_k}_{u_i,p}$ between the cold-start users' mapped distribution and the historical check-in features of all users within the "business circle":

$$sim^{t_k}_{u_i,p}=\cos(\widehat{\beta}^{t_k}_{T,u_i},\beta^{t_k}_{T,u_i}) \qquad (2)$$

where $\widehat{\beta}^{t_k}_{T,u_i}$ represents the cold-start users' mapped topic-

word distribution and $\beta_{T,u_i}^{t_k}$ represents the historical check-in topic-word distribution of all users within the "business circle".

After obtaining the similarity between the cold-start users' mapped distribution and the historical check-in preference of all users within the "business circle", we use the following formula to compute the score of a cold-start user for a POI that has not been checked in:

$$r_{u_i,p} = \sum_{k=1}^{4} \widehat{\theta}_{T,u_i}^{t_k} \cdot \theta_{T,p}^{t_k} \cdot sim_{u_i,p}^{t_k} \quad (3)$$

The top-N POIs are selected to make recommendations for users. Users' interests change over time; a high total score of four quarters represent the user's long-term interest, whereas a low score indicates that the POI is the user's short-term interest.

V. FRAMEWORK ANALYSIS

A. Security analysis

We assume that participants are honest and curious. They will follow the protocol we defined, but they will try to get information from other participants and infer the users' preference. The server is untrusty and may be attacked by third parties or actively reveal users' privacy information. We also assume that cold-start users have historical data in the auxiliary domain and their common users of both auxiliary and target domains are known. Our security goal is guaranteeing that no other participants can obtain users' historical data but the local clients. An attacker cannot infer users' privacy by attacking the train model.

A neural network algorithm [19] is applied to ensure training the encrypted features. Local clients use the homomorphic encryption scheme designed by Li et al. [20] to protect their data. Its security depends on the difficulty of solving a nonlinear system. To encrypt a plaintext m_i, we need to use two parameters q, s and a random number r_i. From a known (m_i, c_i), an attacker needs to solve the nonlinear equation $c_i = s^d (r_i q + m_i) \bmod p$ involving three unknown numbers q, s, r_i and where c_i is the chipertext of m_i. It is an NP-hard problem and thus, achieves data privacy-preservation. Li et al. [20] has verified that the scheme has been greatly improved over other homomorphic encryption schemes in terms of efficiency. Since the mapped feature distribution cannot reflect users' preferences in the auxiliary domain explicitly, this can guarantee the user's privacy on the server side.

B. Algorithm complexity analysis

When a new user checks in, the user is a cold-start user for the POI field. We divide each user's similar user set into four quarters for calculation the long and short term. The division can enhance the degree of long-term and short-term interest reflected by the similar user set.

Algorithm 1 is related to the size of the cold-start users' similar user set. We assume that the size of the similar user set in one quarter is $SIM_{u_i}^{t_k} = \{u_1, u_2, \ldots, u_s\}$, the core of algorithm1 is the LDA model (defined in Section III. B) applied among the four quarters. The algorithm complexity of the LDA model is $O(N_{iter} N_u K)$ where N_{iter} is the model's iterations number, N_u is the total number of all users' purchase records, and K is the number of model topics. Thus, the complexity of the LDA model in our algorithm is $O(4 \bullet N_{iter} N_u K)$. In the encryption algorithm, data storage is distributed. The encryption algorithm is a linear complexity algorithm. Suppose there are m cold-start users in total, the complexity of the encryption algorithm is $O(m \cdot n)$. The complexity of algorithm 1 is the sum of the time of two algorithms: $O(4 \cdot N_{iter} N_u K + m \cdot n)$.

Since the target domain uses offline data stored locally and is not affected by the growth of user scale, we did not analyze it in detail. The server stores only one MLP training model to train the latent feature distribution after encryption which reduces the computation overhead brought by homomorphic encryption algorithm. Firstly, we assume that the ith layer of the neural network is n_i. Since our neural network is fully connected, the time it takes to process a piece of data is $O(n_1 \times n_2 + n_2 \times n_3 + \ldots + n_{i-1} \times n_i) = O(n^2 \cdot (i-1)^2)$. The samples of latent feature mapping are the similar user set of cold-start users and s represents the size of the similar common user set for the cold start user; therefore, the final algorithm complexity is $O(m \cdot s \cdot n^2 \cdot (i-1)^2)$. The server sends the encrypted feature distribution to the POI field and the complexity of the encryption algorithm is $O(m \cdot s \cdot n^2 \cdot (i-1)^2 + m \cdot n)$.

VI. EXPERIMENTAL STUDY

We choose two open datasets for our performance measurement experiments: Movielens and Foursquare, which are widely used in POI recommendation. Movielens contains historical data generated by 138,493 users between 2010 and 2015. After preprocessing, it contains the historical tag data of 46,274 users. At the same time, we use FourSquare to analyze users' check-in preferences, including the historical check-in data generated by 22,743 users between 2012 and 2013. It contains descriptive information about POI, such as coffee, home, medical sites, etc. In our experiments, we use the Movielens dataset as the auxiliary domain data and the Foursquare dataset as the target domain data. The Precision and Recall are used to measure the performance of FL&PP-POI.

$$Precision = \frac{\sum_U |R(u) \bigcap T(u)|}{R(u)} \quad (4)$$

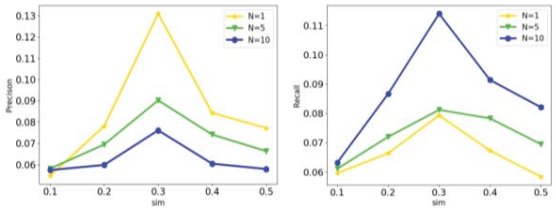
$$Recall = \frac{\sum_U |R(u) \bigcap T(u)|}{T(u)} \quad (5)$$

where $R(u)$ represents the recommended list generated by the recommendation system and $T(u)$ represents the actual POI check-in list.

Since the feasibility of training encrypted user latent features in a neural network [19] and efficiency of implementation of homomorphic encryption [20] have been proved, we focus on testing recommended accuracy. By comparing with two

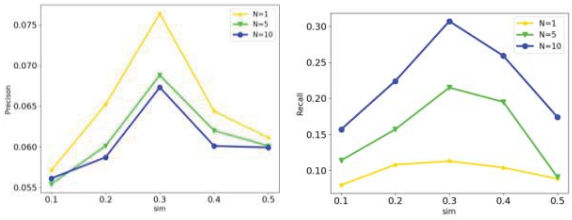
methods, the data fusion method without using machine learning and the method using machine learning but without data fusion, we demonstrate that our proposed framework improves recommended accuracy while protecting user's privacy.

We first analyze the effect of cold-start user' similar users set. We set 70% of the original data as the common user set, and 30% of the common user set as the test set to measure the influence of similarity *sim*, under a different topic value *K*. Fig.1-3 show the influence of *sim* on accuracy (i.e., precision) when *K*, the number of topics in LDA model, is set to 10, 15 and 20 respectively.



(a) Precision (b) Recall

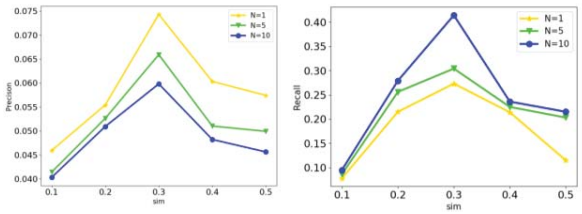
Fig. 1. Recommendation Effectiveness when *K*=10



(a) Precision (b) Recall

Fig. 2. Recommendation Effectiveness when *K*=15

We can see from Figures 1-3 that regardless of the value of *K*, the result of the recall is the highest when *sim* is equal to 0.3. In the case of *N*=10, the variation of accuracy is more stable, and the recalls are generally among the highest. The recalls reflect user check-in results that are ultimately included in the recommendation list. Here, *N* represents the length of the recommendation list.



(a) Precision (b) Recall

Fig. 3. Recommendation Effectiveness when *K*=20

In addition to the influence of similarity on the results, we also measure the influence of common user sets with different densities. The experimental results are shown in Tables 1-3. As seen in Tables 1-3 as the common user set's density increases, the recall increases while the accuracy decreases. Precision and recall are mutually restricted. In order to achieve a better balance, that is ensuring a high recall, the value of accuracy should be improvement. When we take the common user set at the density of 50%, our recall can reach 40% when subject number *K*=20. However, high recalls may lead to accuracy instability. When *K*=15, the precision and recall can reach around 10% at the density of 50%. Thus, we chose the value of *K*=10 in the following experiments.

Table 1. Comparison of Recommendation Effectiveness under 30% density of common users

	Precision			Recall		
	K=10	*K*=15	*K*=20	*K*=10	*K*=15	*K*=20
N@1	0.268	0.0868	0.0948	0.0591	0.193	0.184
N@5	0.201	0.0765	0.0864	0.0699	0.257	0.193
N@10	0.120	0.0585	0.0693	0.0872	0.361	0.223

Table 2. Comparison of Recommendation Effectiveness under 50% density of common users

	Precision			Recall		
	K=10	*K*=15	*K*=20	*K*=10	*K*=15	*K*=20
N@1	0.131	0.0764	0.0743	0.0793	0.113	0.273
N@5	0.0902	0.0688	0.0659	0.0812	0.215	0.304
N@10	0.0761	0.0673	0.0598	0.114	0.307	0.414

Table 3. Comparison of Recommendation Effectiveness under 70% density of common users

	Precision			Recall		
	K=10	*K*=15	*K*=20	*K*=10	*K*=15	*K*=20
N@1	0.0952	0.0620	0.0658	0.0690	0.214	0.308
N@5	0.0737	0.0537	0.0599	0.0944	0.311	0.379
N@10	0.0710	0.0476	0.0478	0.145	0.439	0.441

To further evaluate the performance of our FL&PP-POI framework, four typical relevant models described below are compared in our experiments. The results of precision and recall are shown in Table 4.

Table 4. The Comparison Results with other methods

	Recall			Precision		
	N=1	*N*=5	*N*=10	*N*=1	*N*=5	*N*=10
UCF	0.0340	0.0345	0.0352	0.0338	0.0313	0.0285
LCARS	0.0409	0.0439	0.0489	0.0409	0.0380	0.0353
CAPRF	0.0422	0.0461	0.0511	0.471	0.0460	0.0439
CPC	0.0462	0.0469	0.0520	0.0531	0.0517	0.0490
FL&PP-POI	**0.113**	**0.251**	**0.307**	**0.0764**	**0.0688**	**0.0673**

1)UCF[22]: It is a content-based recommendation system, which adopts the user-based collaborative filtering method. The predicted value of any POI is calculated based on the check-in behavior of similar users. It relies on historical check-in data.

2)LCARS[2]: It is a cross-domain recommendation framework based on LDA model. It improves the recommendation accuracy by fusing the users' topic feature distribution in two domains. They combined the comments in POI domain with historical tag data from users in another domain (Douban).

3)CAPRF[9]: It integrates multiple types of comments into a POI recommendation framework to alleviate data sparsity. It uses matrix factorization technology to analyze users' latent features.

4)CPC[1]: It is a POI recommendation method that applies the convolutional neural network to alleviate the sparsity of POI check-in data. They integrated users' emotional information in the comments.

As seen in Table 4, our method increases the precision and recall rate by 0.37% and 4.9% compared with CPC[1] respectively. It is significantly better than those methods that simply consider comments in POI field. The *precision* can reach 10% when *N*=1. The results are due to us applying cross-domain correlation instead of simply using the comments generated by users in POI field. Meanwhile, in our FL&PP-POI framework, we also use data from e-commerce fields to create a more accurate description for user patterns. More importantly, our data is encrypted. Although the correlation between two different domains is considered in [2], machine learning and privacy protection are not included. Compared to [2], our FL&PP-POI framework improves the recommendation accuracy and recall by 0.9% and 5.3% respectively, while protecting user's privacy at same time.

VII. CONCLUSION

Presently, in the cross-domain POI recommendation, there are few works considering the risks in data privacy brought by cross-domain data correlation. In this paper, we propose an efficient POI recommendation framework, FL&PP-POI, with cross-domain correlation and privacy protection features. We take e-commerce data as an example when conducting a cross-domain POI recommendation. The accuracy of target user patterns is improved while the influence of geographical location and time factors on user check-in preference are also considered. The federated learning mechanism is applied when storing users' original data locally. We also adopt homomorphic encryption technology to further protect the users' privacy in the necessary data processing and exchange. Experimental results have shown that FL&PP-POI is advantageous over many relevant techniques. In particular, it effectively improves recommendation accuracy with a moderate density of common user set, meaning that it reduces computation overhead and the possibility of privacy leakage.

ACKNOWLEDGEMENT

This work is in part supported by the National Natural Science Foundation of China (No.61662008), the Guangxi Natural Science Foundation (Nos. 2020GXNSFAA297075, 2018JJA170082), the Guangxi "Bagui Scholar" Teams for Innovation and Research Project, the Guangxi Collaborative Innovation Center of Multi-source Information Integration and Intelligent Processing, the Guangxi Talent Highland Project of Big Data Intelligence and Application, the Research Fund of Guangxi Key Lab of Multi-source Information Mining & Security (No. 19-A-02-02) and the PostGraduate Education Innovation Project of Guangxi Normal University under grant JXXYYJSCXXM-006. This work is also in part supported by the National Science Foundation (NSF) Grant 1921576.

REFERENCES

[1] X.S. Ning, F.A Liu, Q.Q Wang, X.H Zhao, T.L Li. "Content-aware point-of-interest recommendation based on convolutional neural network," Applied Intelligence. 2018,49(3):858-871.

[2] H.Z. Yin, Y.Z. Sun, B. Cui, Z.T Hu, L. Chen. "LCARS: a location-content-aware recommender system," in ACM Sigkdd International Conference on Knowledge Discovery & Data Mining. Chicago, ACM, 2013, pp. 221-229.

[3] G.Q. Liao, S. Jiang, Z.H Zhou, C.X Wan, X.P Liu. "POI Recommendation of Location-Based Social Networks Using Tensor Factorization," in International Conference on Mobile Data Management. Denmark, IEEE, 2018, pp. 116-124.

[4] D.C. Li, Z.G Gong, D.F Zhang. "A Common Topic Transfer Learning Model for Crossing City POI Recommendations," IEEE Transactions on Cybernetics. 2018,49(12):4282-4295.

[5] J.G. Ding, G.H. Yu, Y. Li, D.P Jin, H. Gao. "Learning from Hometown and Current City: Cross-city POI Recommendation via Interest Drift and Transfer Learning," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2019,3(4):1-28.

[6] J. Zeng, qX. He, F. Li, Y.B. Wu. "A recommendation algorithm for point of interest using time-based collaborative filtering," International Journal of Information Technology and Management. 2020,19(4): 347-357.

[7] Y.D. Seo, Y.S. Cho. "Point of interest recommendations based on the anchoring effect in location-based social network services," Expert Systems with Applications. 2021,164:114018.

[8] B. Chang, Y.G. Park, D. Park, S. Kim, J. Kang. "Content-Aware Hierarchical Point-of-Interest Embedding Model for Successive POI Recommendation," in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, IJCAI, 2018, pp. 3301-3307.

[9] H.J. Gao, J.L. Tang, X. Hu, H. Liu. "Content-Aware Point of Interest Recommendation on Location-Based Social Networks," in Twenty-ninth AAAI Conference on Artificial Intelligence. Austin, Texas, USA, AAAI Press, 2015, pp. 1721-1727.

[10] D. D'Agostino, F. Gasparetti, A. Micarelli, G. Sansonetti. "A Social Context-Aware Recommender of Itineraries Between Relevant Points of Interest," in International Conference on Human-Computer Interaction. Toronto, Canada, Springer,2016, pp. 354-359.

[11] Z.Y. Zhang, Y. Liu, Z.J. Zhang, B. Shen. "Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation," World Wide Web. 2018,22(3): 1135-1150.

[12] Y. Wen, J.S. Zhang, Q.T. Zeng, X. Chen, F. Zhang. "Loc2Vec-Based Cluster-Level Transition Behavior Mining for Successive POI Recommendation," IEEE Access. 2019,7:109311-109319.

[13] X.W. Wang, H. Yang, K. Lim. "Privacy-Preserving POI Recommendation Using Nonnegative Matrix Factorization," in IEEE Symposium on Privacy-Aware Computing. DC, USA, IEEE, 2018, pp. 117-118.

[14] X.X. Li, P.P. Sun, Y. Bai, L.E. Wang. "M-generalization for multipurpose transactional data publication," Frontiers of Computer Science.2018,12(6):1241-1254.

[15] L.E Wang, X.X Li. "A graph-based multifold model for anonymizing data with attributes of multiple types," Computers & Security. 2018,72C:122-135.

[16] L. Kuang, S.M. Tu, Y.Q. Zhang, X.X. Yang. "Providing privacy preserving in next POI recommendation for Mobile edge computing," Journal of Cloud Computing.2020,9(1).

[17] W.Q. Wang, A. Liu, Z.X. Li, X.L. Zhang, Q. Li, X.F. Zhou. "Protecting multi-party privacy in location-aware social point-of-interest recommendation," World Wide Web.2018,22(2): 863-883.

[18] C.C. Chen, Z.Q. Liu, P.L. Zhao, J. Zhou, X.L. Li "Privacy Preserving Point-of-Interest Recommendation Using Decentralized Matrix Factorization," in Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, Louisiana, USA, AAAI Press,2018, pp. 257-264.

[19] Q.Z. Wang, L. Gao. "Neural Network for Processing privacy-protected Data," Journal of Cryptologic Research. 2019,6(2):258-268.

[20] L.C. Li, R.X. Lu, K.R. Choo, A. Datta, J. Shao. "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases.2016,11(8):1847-1861.

[21] X.H. Wang, Z.H. Peng, S.Z. Wang, et al. "CDLFM: cross-domain recommendation for cold-start users via latent feature mapping," Knowle dge and Information Systems. 2020,62(5):1723-1750.

[22] D.Q. Zhou, B. Wang, S.M. Rahimi, X. Wang. "A Study of Recommending Locations on Location-Based Social Network by Collaborative Filtering," in Advances in Artificial Intelligence. Toronto, ON, Canada, Springer, 2012, pp. 255-266.