Deep Potential Geo-Social Relationship Mining for Point-of-interest Recommendation

ZHENGGO PAN1, LIN CUI1, XIAOYIN WU1, ZHIWEI ZHANG1, XIANWEI LI1, GUOLONG CHEN1.

1School of Informatics and Engineering, Suzhou University, Suzhou 234000, P.R.China
Corresponding author: Lin Cui (e-mail: jsjxcuilin@126.com).

This work was supported in part by the National Natural Science Foundation of China No.61702355, in part by Natural Science Foundation of Anhui No.1908085qf283, in part by the Daze Scholar Project of Suzhou University No.2018szxydzx01, in part by the Key Natural Science Project of Anhui Provincial Education Department No.kj2018a0448 and No.kj2019a0668, in part by Anhui Provincial Outstanding Young Talent Support Project No.gxyq2017093, and in part by Scientific and Technological Projects of Suzhou University No.sz2017gg39 and No.sz2018gg03, in part by Key Discipline of Computer Science and Technology at Suzhou University No.2019xjdzxk1.

ABSTRACT Point-of-Interest (POI) recommendation has been an important research topic in data mining and the popularity of Location Based Social Networks (LBSNs) has significantly contributed to POI recommendation. Existing POI recommendation models mostly adopt various explicit social relationships under geographical space. The implicit relationships among users under a certain geographical region are rarely taken into account, though they have major influence on user behaviors. Due to the above limitations, we were motivated to introduce an innovative Deep Potential Geo-Social Relationship mining model for POI Recommendation (DPGSR-PR). The proposed DPGSR-PR performs the personalization of geographical features, towards the determination of user check-in behaviors and choices in specific domains, which is achieved by estimating and considering kernel density. Moreover, user preferences and personalized geo-social influence are incorporated into a geo-social recommendation framework under a holistic view. Specifically, the deep potential geo-social relationships include the Explicit-Implicit User Geo-Social Relationships between users (EIU-GSR) and the Implicit Common check-in POI-based Geo-Social Relationships (ICP-GSR). The estimation of Kernel Density and the two-hop random walk approach are employed in an effort to mine the EIU-GSR. The ICP-GSR are discovered by specifically determining and considering the Jaccard similarity coefficient and kernel density estimation. Due to the fact that the role of both EIU-GSR and ICP-GSR as regularization terms is quite significant, we used their combined impact to obtain a unique recommendation model that employs matrix factorization. The proposed DPGSR-PR was tested on two datasets, which has proven that DPGSR-PR outperforms other well-established models.

INDEX TERMS Geo-social relationship, Point-of-interest recommendation, Check-in activities

I. INTRODUCTION

THE huge technological advances in the fields of mobile and wireless communications make the rapid progress of location acquisition technologies, and also have offered an immense popularity and power to LBSNs, like Facebook, Gowalla and Foursquare and so on. The use of such media has become a part of every-day life for millions of users globally [1-7]. Through utilizing LBSNs, the users can exchange views and share their stories, thoughts and experiences (good or bad). The concept ‘checked-in’ is widely used in the literature to describe user preferences regarding various locations in the globe and POI is adopted to describe locations that are checked-in. Some researches on POI recommendation [8-12] conclude that it is very likely for people to be interested in and to search for POIs that are located close to the ones that they have already visited, which leads to the spatial clustering of POIs based on user choices. Other well-known researches enhance the efficiency and performance of the recommendations by exploiting social friendships [13-15], which is performed by evaluating the similarities among all socially connected users and utilizing them as regularization terms to constrain matrix factorization, and experiments on real-world datasets have validated that focusing on social information is effective.

Consequently, our research team has been highly motivated to mine the implicit social relationships in specific
Influence of deep potential geo-social relationships on POI recommendation. The lower level denotes the explicit-implicit geo-social relationships and the top level presents POI recommendation performed by users who have the same check-ins with the target user in similar geographical regions.

The proposed DPGSR-PR model is a novel POI recommendation approach, which is the first model that simultaneously considers two deep potential geo-social relationships, namely EIU-GSR and ICP-GSR. The two deep potential geo-social relationships are regarded as regularization terms, which are used to constrain matrix factorization and to enhance recommendation accuracy.

The first deep potential geo-social relationships EIU-GR mined the explicit and implicit user relationships by adopting the proposed two-hop random walk algorithm and kernel density estimation. The second geo-social relationships ICP-GSR include the implicit common check-in POI-based geo-social similarity between users, which are computed using Jaccard similarity coefficient and kernel density estimation. Moreover, the top-N POI result is recommended to the target user by using a combination of EIU-GR and ICP-GSR in a hybrid manner.

Several experiments have been performed in an effort to evaluate DPGSR-PR on two datasets from Foursquare and Gowalla. DPGSR-PR is verified compared with other five well-known baseline models, which proves that our proposed DPGSR-PR has the best performance. The metrics used are precision rate, recall rate and NDCG@N that is normalized discounted cumulative gain.

The remaining is presented as below. Section II lists related work. Section III introduces some basic principles, which clearly defines and states the case of problems. Section IV discusses the actual use of the kernel density estimation and geometric areas. An intuitive case has been used to demonstrate the actual significance level as depicted in FIGURE 1.
its role in modeling the personalized geographical impacts on user behaviors. Moreover, the DPGSR-PR integrating user preference, social and personalized geographical factors is introduced. Section V clearly explains the optimization of DPGSR-PR by employing the Stochastic Gradient Descent (SGD) method and discusses the time complexity of DPGSR-PR. Section VI is dedicated to the performance evaluation of DPGSR-PR and the analysis of experimental results. Finally, section VII summarizes our research and presents future work that would potentially further optimize the whole effort.

II. RELATED WORK

The classification of recommendation technologies and POI Recommendation based on social and geographical factors are introduced in this section, then we list the connection to prior work.

A. CLASSIFICATION OF RECOMMENDATION TECHNOLOGIES

The general recommendation approaches are divided into two classes that are the content-based ones [16] and collaborative filtering ones [17]. The content-based recommendation mainly recommends the items to the user according to the relevance of content information of recommended items and user profiles. As the content-based recommendation method cannot flexibly combine verified information related with users, it is rarely used for POI recommendation. The collaborative filtering methods generally include memory-based and model-based recommendations, among which, the former contains user-based and item-based recommendation methods. Ye et al. first adopted the memory-based method to perform POI recommendation [18]. However, as the memory-based methods usually face with the cold start and data sparsity, etc., some advanced POI recommendation methods have been proposed by improving the traditional memory-based method [19-21]. The model-based recommendations have been widely used in POI recommendation because of its advantages [22, 23], which fits the original rating matrix to be a reduced model and its sparsity is also decreased by the probability distribution or the low dimensional matrix, then this model is adopted to forecast the score of not-rated items.

Matrix factorization is a widely used model-based approach [24, 25], which describes that user preferences are determined by a group of latent representations and are also used to properly model users and latent items (e.g. POI, products). Zhao et al. [24] introduced a tensor factorization based on Bayesian probability to mine social factors for recommendation, this methodology mainly focuses on mining the heterogeneous relationships among users, places and time. Hu et al. [25] proposed a model to perform geographical rating prediction, which employs an algorithm taking into account the intrinsic features of business and extrinsic features of their geographical neighbors as well.

B. SOCIAL AND GEOGRAPHICAL-BASED POI RECOMMENDATION

POI recommendation can fulfill the specific individual requirements of users who wish to travel to new areas, and can also help LBSNs companies to raise their revenue. Offering accurate location services is a key issue as it can provide the user with significant aid and contribute to the development of companies. Our paper is dealing mainly with social POI recommendation and geographical POI recommendation.

This paragraph describes social POI recommendation. POI recommendation approaches have contributed towards enhancing the recommendation effectiveness by taking into account social influence [6, 15, 18]. An interesting approach using friend-based collaborative filtering for POI recommendation was introduced by Ye et al. [18]. It is an algorithm that considers the common visited check-ins of friends. The disadvantage of this method is that it only takes into account friend preferences. In 2011, Ye et al. [6] introduced the social influence weight among couples of friends into recommendation, which is founded on the existing social connections and on the level of similarity between the check-in preferences of friends. Cui et al. [15] presented a kind of social recommendation based on the deeper membership and friendship, which presented that social recommendation is very important for POI recommendation.

Geo-based POI recommendation is discussed in this paragraph. POI recommendation systems are distinguished from the usual recommendation systems by considering geographical influence. It is a fact that they have strong influence on the traveling profile of users. Tobler [26] introduced the first geography law, which declares that the similarity of things closer to each other is greater than that of things farther away, that is to say, nearby places would be more likely to be arrived compared to the distant ones. Recent and well established researches have proven that geographical contiguity greatly affects user check-in behaviors [12, 27]. Cheng et al. [12] suggested that people prefer to travel to locations around specific centers. He also studied this problem statistically and supported the argument that the visited locations present Gaussian distribution, based on which, they introduced the multi-center Gaussian model to analyze the geographical factor of POI and they have integrated it with the matrix factorization approach to perform a hybrid POI recommendation. It has been noticed by Liu et al. [27] that people tend to get to locations in short distance from their homes or their work places, and they point out that the distance between two positions checked-in by the same person is close to power-law distribution. The limitation of the above studies is that, for all users, they adopt the unified distance distribution to model the location impacts.

C. CONNECTION TO PRIOR WORK

Obviously, user preferences, social relations and geographic characteristics are strongly influencing the user check-in behaviors. Though there exist many POI recommendation researches that take into account the above factors, our
approach proposed herein, is only one that simultaneously considers the explicit-implicit geo-social relationships between users and the implicit common check-in POI geo-social relationships of the destination area. The deep potential geo-social relationships are incorporated through matrix factorization in our DPGSR-PR model.

III. PRELIMINARIES

This chapter presents relative data structures, which are used to store the historical check-in data related to POIs and all other involved factors (e.g., POI social links, geographical longitude and latitude coordinates). Moreover, the research problem is clearly presented and defined.

In our research, $U = \{u_1, u_2, \ldots u_m\}$ presents user set, and $P = \{p_1, p_2, \ldots p_n\}$ is POI set. If user $u_i$ has checked-in POI $p_j$, we stipulate $r_{ij} \neq 0$, otherwise $r_{ij} = 0$. In addition, $F$ represents the friend set and $F_i$ denotes all friend set of user $u_i$. In this case, POI recommendation is transformed to the recommendation of new POIs to user $u_i$ based on user check-in behaviors. We present several notations that have been employed in our research in Table 1.

In the next paragraphs, we introduce five concepts related to the DPGSR-PR approach.

**Definition 1. POI**: POI usually refers to a solely identified event or a venue. We stipulate that it has two basic characteristics namely an identifier represented as $p$ and a location (corresponding to longitude and latitude coordinates) denoted as $l$.

**Definition 2. POI recommendation**: POI recommendation refers to recommending new POIs to the target user, it should be clarified that these POIs are not in POI set $P_i$ that the target user has visited.

**Definition 3. Check-in activity**: The tuple $(u, p, l_v, l_w)$ is used to denote user check-in activity, among which, the tuple parameters are explained as follows: user $u$ travels at POI $p$ in location $l_v$, in which, $l_v$ is the coordinate of the corresponding longitude and $l_w$ is the coordinate of the corresponding latitude.

**Definition 4. Social link matrix**: Social link matrix $S_{|U| \times |U|}$ comprises of social links among users. If two users $u_i$ and $u_j$ are connected with a social link, then $S_{u_i, u_j} = 1$ and $u_j \in F_i$.

**Definition 5. Latent factor matrix**: The notation $U \in R^{m \times k}$ is adopted for the user latent factor matrix, which contains the specific user check-in preferences based on some latent parameters. $P \in R^{k \times n}$ is the POI latent factor matrix and it comprises of latent factors, in which, we use $k \ll (m, n)$ to denote the dimension of latent factors.

### A. PROBLEM STATEMENT

This research was motivated to find the individual geographic influence of places and the individual geographic influence is regarded as a unique distance distribution for every user. Moreover, the personalization geographic influence of locations was integrated with user preferences and social influence. Finally, DPGSR-PR produces a POI recommendation list for the target user $u_i$, which is achieved by applying a mining process based on EIU-GR and ICP-GSR.

IV. METHODOLOGY

In this section, we discuss all aspects and principals related to the proposed DPGSR-PR. We first present the explicit-implicit user geo-social relationship mining process EIU-GR. Then, we describe the potential geo-social relationship mining process ICP-GSR. Subsequently, we adopt the matrix factorization to combine these two latent relationships and produce DPGSR-PR.

**A. THE EXPLICIT-IMPLICIT USER GEO-SOCIAL RELATIONSHIP MINING**

Personalized geographical influence of locations is a very important process, which affects the specific user behavior. The distance of personalized distribution among two locations checked-in by the same user is deduced by adopting kernel density estimation, which shows that most users always check-in places in short distances. As a matter of fact, the desire and determination to visit an area is inversely depending on the distance from the current location. Hence, we adapt the kernel destination estimation of distance to denote the user willingness from one place to another [28, 29]. We first compute the distance between locations as the following equation (1).

$$d_{x_0} = \text{distance}(l_{x}, l_{o}), \forall l_{o} \in L_i$$ (1)
where \(d_{x_o}\) represents the distance between \(l_x\) and \(l_{o}\), and \(l_{o}\) belongs to the visited POI set \(L_t\) of the target user. \(d_{x_o}\) is employed to compute the probability using equation (2),

\[
\hat{f}(d_{x_o}) = \frac{1}{|D|h} \sum_{d^e \in D} K\left(\frac{d_{x_o} - d^e}{h}\right)
\]

where \(D\) is derived from the distribution under density \(f\) for the case of a specific user. \(K(\cdot)\) denotes a kernel function and \(h\) is a path distance attenuation threshold, which is called the bandwidth. This research employs the most widely used normal kernel as shown in equation (3).

\[
K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}
\]

We must specify that the optimal bandwidth \(h\) is defined as the following equation (4) [25].

\[
h = \left(\frac{\lambda h^5}{3n}\right)^{1/5} \approx 1.06\sigma n^{-1/5}
\]

where \(\sigma\) denotes the sample standard deviation in \(D\), \(n\) is the size of POI set whose path distance from the position of the target user is less than or equal to \(h\).

The algorithmic steps are as follows: we first determine the distance distribution according to the kernel destination estimation, in the second step, equation (2) is used to develop an approach towards the probability estimation of user \(u_i\) to visit a new location \(l_x\) considering the locations \(L_t\) that the same user has visited. Subsequently, in the third step, we compute the probability of \(u_i\) to visit a new place \(l_x\) based on the mean probability as the following equation (5).

\[
p(l_x|L_t) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(d_{x_o})
\]

The latent relationships are mined for obtaining users who have the link relationship with the target user, and the two-hop random walk method is adopted to determine the link relationship similarities among users. The link prediction employs the random walk process, which specifies the movement from a specific user to any neighbor nodes [30, 31]. The implicit relation between \(u_i\) and \(u_t\) is decided using the following equation (6).

\[
score(e) = \sum_{p \in paths(i, t)} \left[ \prod_{e' \in edges(p)} p(e') \right] \quad \text{s.t.} \, |edges(p)| \leq 3
\]

in equation (6), \(score(e)\) stands for the rating of an unknown link that exists between \(u_i\) and \(u_t\). The value of \(score(e)\) corresponds to the implicit correlation degree between user \(u_i\) and user \(u_t\). Moreover, \(paths(i, t)\) is all potential path set from \(u_i\) to \(u_t\), whereas the \(edges(p)\) is a set including all edges in path \(p\), and \(p(e')\) is the probability of edge \(e'\) to be chosen in the process of random walk. It is important to notice that \(p(e')\) is proportional to the weights assigned to edge \(e'\). Finally, the higher the sum of weights assigned to the edges related with the initial node of \(e'\), the smaller \(p(e')\) as shown in equation (7).

\[
p(e') = \frac{w(e')}{\sum_{(x')} w(x')}
\]

in equation (7), the weights of edge \(e'\) and the initial vertex of edge \(e'\) are represented by \(w(e')\) and \(x'\), respectively. In addition, \(\langle x'\rangle\) is the adjacent edge set to edge \(x'\).

**Algorithm 1** The First Deep Potential Geo-social Relationship Mining Algorithm EIU-GSR

**Input:**
- Original user-POI check-in matrix \(R \in R^{m \times n}\), the target user \(u_t\), the observed check-ins \(R_{u,p}\), social link matrix \(S\), the initial walking probability vector \(P^0\).

**Output:**
- \(w_{it}\): Similarity based on explicit-implicit link relationships between user \(u_i\) and user \(u_t\);
- 1: Compute the average core distance \(d^e\) of the target user \(u_i\);
- 2: Perform kernel density estimation using equation (1), (2), (3), (4) and (5);
- 3: Obtain the probability \(p(l_x|L_t)\) of \(u_i\) visiting a new location \(L_t\);
- 4: Execute two-hop random walk algorithm in the region where the target user \(u_i\) can arrive using equation (6) and (7);
- 5: Until convergence
- 6: Obtain the social similarity between user \(u_i\) and other users based on EIU-GSR by equation (6);
- 7: Calculate the geo-social similarity \(w_{it}\) using equation (8);
- 8: Return \(w_{it}\)

The final obtained graph is undirected and it consists of the corresponding sets of relative users. The graph complexity is high due to the actual high complexity of user set. Thus, when our research processes random walk approximately, the paths with the number of edges less than or equal to 3 are considered, which is the basic characteristic of our proposed two-hop style as shown in (6). The preference of explicit-implicit friends of the target user on POIs, and their personalized mobility patterns are inferred by using kernel density estimation in equation (5), which are fused in order to obtain the probability \(u_i\) checking-in at a POI \(l_x\). Thus, the similarity between users \(u_i\) and \(u_t\) is decided by their social relationships and geographical distance (see the following equation (8)).

\[
w_{it} = p(l_x|L_t) \cdot score(e)
\]
B. THE IMPLICIT COMMON CHECK-IN POI-BASED GEO-SOCIAL RELATIONSHIP MINING

In terms of POI $p_i$ if user $u_t$ and $u_c$ both visited $p_i$, they share similar interests. Since every user’s interest can be represented by their checked-in POIs, therefore, Jaccard similarity coefficient is employed to estimate the similarity between users $u_t$ and $u_c$ (See (9)).

$$Q_{ic} = \frac{|E(u_t) \cap E(u_c)|}{|E(u_t) \cup E(u_c)|}$$

(9)

where $Q_{ic}$ represents the interest similarity between users $E(u_t)$ includes POIs which have been checked-in by user $u_t$, and $E(u_c)$ is related to POIs checked-in by user $u_c$.

The second implicit common check-in POI-based geo-social relationship mining performs further fusion of all the POI check-ins as given by equation (9). The probability is obtained by the personalized geographical influence of locations described in equation (5), and the geo-social similarity $s_{ic}$ among user $u_t$ and user $u_c$ based on ICP-GSR is obtained. The geo-social similarity $s_{ic}$ is expressed as follows:

$$s_{ic} = p(l_x | l_i) \cdot Q_{ic}$$

(10)

The second deep potential geo-social relationship mining algorithm ICP-GSR is presented by Algorithm 2.

Algorithm 2 The Second Deep Potential Geo-Social Relationship Mining Algorithm ICP-GSR

Input: Original user-POI check-in matrix $R \in R^{m \times n}$, the target user $u_t$, the observed check-ins $R_{u_t,p}$

Output: $s_{ic}$: Similarity between user $u_t$ and user $u_c$ based on the common visited check-in relationships in a certain geographical region

1. Compute the average core distance $d'$ of the target user $u_t$;
2. Perform kernel density estimation using equation (1), (2), (3), (4) and (5);
3. Obtain the probability $p(l_x | l_i)$ which denotes that $u_t$ visits a new location $l_x$;
4. Execute Jaccard similarity coefficient in the region where $u_t$ can arrive using equation (9);
5. Obtain the geo-social similarity $s_{ic}$ between $u_t$ and $u_c$ using equation (10);
6. Return: $s_{ic}$

C. THE UNIFIED DPGSR-PR MODEL

In order to mine the deep potential geo-social relationships, we combine EIU-SGR and ICP-SGR, and propose a deep potential geo-social relationship mining-based POI recommendation model (DPGSR-PR). DPGSR-PR is based on matrix factorization, which comprehensively combines user preference, social friendships and geographical information. The description of DPGSR-PR is shown in the following paragraphs.
problem. The two implicit geo-social relationships from EIU-GSR and ICP-GSR are regarded as regularization for the purpose of constraining matrix factorization and achieving more accurate predictions in \( R \).

V. OPTIMIZATION AND TIME COMPLEXITY ANALYSIS OF DPGSR-PR

The optimization process of DPGSR-PR is discussed firstly and we present the corresponding model parameters using the Stochastic Gradient Descent (SGD) [33]. Then, the time complexity of DPGSR-PR is discussed and analyzed.

A. OPTIMIZATION OF DPGSR-PR

In this paper, SGD worked as an optimizer to optimize the objective function. SGD performs random canning on all training dataset and updates model parameters of objective function for the POI entry of each user. The equation (14) below performs each update.

\[
\Lambda \leftarrow \Lambda - \xi \cdot \frac{\partial F(\Lambda)}{\partial \Lambda} \quad (14)
\]

it should be clarified that \( \xi \) denotes the learning rate, \( \Lambda \) includes all adopted parameters, and the partial derivative \( \frac{\partial F(\Lambda)}{\partial \Lambda} \) represents the objective function, which is shown in equation (13).

The gradient estimation of \( \frac{\partial F(U_i, P_j, S, SC)}{\partial U_i} \) with respect to \( U_i \) is performed according to equation (15). \( U_k \) indicates a node set, while the target nodes are fixed and the local minimum of objective function can be computed by using gradient descent strategy (see equation (15) and (16)).

\[
\frac{\partial F}{\partial U_i} = \sum_{j=1}^{N} I_{ij}(U_i^T P_j - R_{ij})P_j + \beta U_i + \beta \sum_{t \in S} w_{it}(U_i - U_t) + \gamma \sum_{c \in SC} s_{ic}(U_i - U_c)
\]

Therefore, \( U_i \) is updated as

\[
U_i \leftarrow U_i - \xi \cdot \frac{\partial F(U_i, P_j, S, SC)}{\partial U_i} \quad (16)
\]

The gradient of \( \frac{\partial F(U_i, P_j, S, SC)}{\partial P_j} \) with respect to \( P_j \) is given in equation (17)

\[
\frac{\partial F}{\partial P_j} = \sum_{i=1}^{M} I_{ij}(U_i^T P_j - R_{ij})U_i + \alpha P_j
\]

and \( P_j \) is updated as in equation (18)

\[
P_j \leftarrow P_j - \xi \cdot \frac{\partial F(U_i, P_j, S, SC)}{\partial P_j} \quad (18)
\]

After the estimation of gradients \( \frac{\partial F}{\partial U_i} \) and \( \frac{\partial F}{\partial P_j} \), the parameters can be updated by using equation (18). The following pseudo-code clearly introduces the main idea of DPGSR-PR. (See Algorithm 3).

Algorithm 3 The Proposed Algorithm of DPGSR-PR

**Input:**

Original user-POI check-in matrix \( R \in R^{m \times n} \), the dimension \( K \) of latent factors, parameters \( \beta \) and \( \gamma \), the learning rate \( \xi \), the observed check-ins \( R_{u,p} \) and recommendation size \( N \).

**Output:**

Top-N recommended POI list

1: Establish indicator matrix \( I \);
2: Initialize \( U \in R^{m \times k} \) and \( P \in R^{k \times n} \);
3: for each POI \( p \) in POI set \( P \) in the region where \( u_i \) can arrive do
4: Compute \( \frac{\partial F}{\partial P} \) by equation (15);
5: Compute \( \frac{\partial F}{\partial U} \) by equation (16);
10: Compute \( \frac{\partial F}{\partial U} \) by equation (17);
11: Update \( P \) by equation (18);
12: until Convergence
13: \( R_{u,p} = U^T P \);
14: Return: Top-N POI list based on recommendation scores

B. TIME COMPLEXITY ANALYSIS

The time complexity of DPGSR-PR is discussed in this subsection. The overall complexity is defined by the matrix factorization and performing two sub algorithms EIU-GSR and ICP-GSR. In real experiments, because the two sub algorithms EIU-GSR and ICP-GSR are executed offline, the complexity and computational overhead of these two sub-algorithms are not considered. Hence, the time complexity of DPGSR-PR is mainly reflected by calculating the cost of matrix factorization online. In the check-in matrix \( R \), the average observable number of scores per user is \( n_1 \), and the average observable number of check-ins per POI is \( n_2 \). During the iteration of SGD, for all users, the computational complexity of \( \frac{\partial F}{\partial U} \) is \( O(Mn_1K) \), and for all POIs, the computational complexity of \( \frac{\partial F}{\partial P} \) is \( O(Nn_2K) \). Moreover, \( K \) is the dimension of latent factors, and the time complexity of matrix factorization is \( O(Mn_1K + Nn_2K) \).

VI. EXPERIMENTAL RESULTS AND DISCUSSION

For the purpose of validating the accuracy of DPGSR-PR, we made experiments on two real-world datasets. The following two important questions were answered properly: (1) How does the dimension \( K \) of latent factors and regularized parameters \( \beta \) and \( \gamma \) affect the actual performance of DPGSR-PR; (2) How does DPGSR-PR perform compared to the well-established existing methods.

Initially, dataset information and experimental settings are clearly presented. Moreover, evaluation metrics and comparative methods are discussed. In the next step, the proposed DPGSR-PR model is compared with the established-existing...
POI recommendation methods under four different values of $N$. And in the third step, several manipulations are performed to calculate the optimal value of $K$. Finally, experiments are conducted to establish the optimal value of $\beta$ and $\gamma$.

### A. EXPERIMENTAL DATASETS

The used real-world datasets are crawled from Foursquare and Gowalla, respectively. Foursquare as the most widely used online LBSNs collected from Twitter and it was checked-in from January 2011 to July 2011 [3]. Gowalla dataset was obtained from Cheng et al. [12]. However, Gowalla is also aiming to help users check-in at locations by utilizing their mobile devices. The collected Gowalla dataset was checked-in from February 2009 to September 2011. The check-in records in these two datasets contain users’ ID, users’ check-in locations, users’ social relationships and location details. TABLE 2 clearly presents the statistics of datasets.

#### TABLE 2. Statistics of Foursquare and Gowalla

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Foursquare</th>
<th>Gowalla</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of users</td>
<td>11,326</td>
<td>53,944</td>
</tr>
<tr>
<td>No. of POIs</td>
<td>182,968</td>
<td>367,149</td>
</tr>
<tr>
<td>No. of check-ins</td>
<td>1,385,223</td>
<td>4,128,714</td>
</tr>
<tr>
<td>No. of social links</td>
<td>47,164</td>
<td>306,958</td>
</tr>
<tr>
<td>User-location matrix density</td>
<td>0.00067</td>
<td>0.00021</td>
</tr>
</tbody>
</table>

### B. EXPERIMENTAL SETTING

There are several parameters which may affect the results of DPGSR-PR. In the experimental settings that were performed, the dimension $K$ of latent factor is chosen from the set {10, 20, 30, 40, 50, 60, 70, 80}, where 40 is the optimal value. Moreover, the regularization parameters $\beta$ and $\gamma$ are chosen from the set {0, 0.01, 0.05, 0.1, 0.5, 1} and the acceptable combination of $\beta = 0.05$ and $\gamma = 0.05$ is adopted for two datasets. Details of parameter selection can refer to the parameter sensitivity analysis in Subsections $F$ and $G$.

The latent factor are learned by the SGD method. The initial learning rate is equal to $\xi = 0.001$ and parameter $\alpha$ is also set to be 0.001. The same parameters are utilized in all methods to perform a reliable comparison with the baselines. Additionally, the datasets were divided randomly into train set and test set according to 8:2 ratio, respectively. The train set learns the recommendation model, and the test set can prove the level of convergence and show the model generalization ability.

### C. EVALUATION METRICS

The key issue of our research is the effective prediction of a POI recommendation list for the target user. Three evaluation metrics precision rate $P@N$, recall rate $R@N$ [34] and $\text{NDCG}@N$ [35, 36] were utilized to check and to validate the effectiveness of DPGSR-PR, their definitions are shown in equation (19) and (20), respectively:

$$P@N = \frac{\#\text{TestSetHits}}{\#\text{Recommendations}} = \frac{|\text{test} \cap \text{topN}|}{|\text{topN}|}$$

$$R@N = \frac{\#\text{TestSetHits}}{\#\text{Recommendations}} = \frac{|\text{test} \cap \text{topN}|}{|\text{test}|}$$

where $\#\text{TestSetHits}$ are POI set which were checked-in by users in the test set, $P@N$ is adopted to check up the ranking problems of POI recommendation and measures the ratios of the recommended POIs that are truly checked-in in the test set.

$\text{NDCG}@N$ evaluates the recommendation performance according to the graded relevance of recommended items. The evaluation ranges are from 0.0 to 1.0, among which, 1.0 represents the best POI ranking. $\text{NDCG}@N$ is defined as shown in equation (21).

$$\text{NDCG}@N = Z_N \sum_{j=1}^{N} \frac{(2^{r_j} - 1)}{\log(1+j)}$$

where $N$ is the maximum value of recommended entities, $r_j$ equals to 1 if POI at position $j$ is recommended, otherwise, it is 0. $Z_N$ is used for the normalization.

### D. BASELINE METHODS

DPGSR-PR is compared with the following five well-established recommendation approaches, which are presented in the following paragraphs. The aim is to validate the effectiveness of DPGSR-PR and its personalized ranking quality.

1. **BasicMF** [32]. The main idea of BasicMF is based on equation (11). Only the user preference is considered, however, other auxiliary information such as geographical or social data is not taken into account.

2. **Biased MF** [25]. Biased MF is a matrix factorization model with user and POI biases, and it is widely utilized as a baseline for comparing the recommendation performance.

3. **PMF** [37]. In PMF, the rating is calculated by the dot product of user-specific and item-specific features. Latent representations are learned and the rating conversion is performed by using a logistic function, which is done to avoid the data bias.

4. **GeoCF** [6]. GeoCF integrates the geographical influence with the user-based collaborative filtering by adopting a power-law probability. The user preference and geographical influence are linearly combined to construct an unified POI recommendation framework.

5. **CoRe** [29]. CoRe proposed a combined recommendation integrating geographical influence and social influence, which utilizes an estimation approach to predict the probability of a user visiting a new place, then based on their social friendship and location, the similarities among users are computed.

---

1. https://foursquare.com/
2. https://gowalla.com/
E. METHOD COMPARISON ON $P@N$, $R@N$ AND $NDCG@N$

Experimental results are made comparable by employing the metrics $P@N$, $R@N$ and $NDCG@N$. The first step was to determine how $P@N$, $R@N$ and $NDCG@N$ change according to the length of recommendation lists, (e.g., the top-N). During the experiments, $N$ was setted the values of 5, 10, 15 and 20, whereas the latent dimension $K$ was equal to 40. Experimental comparisons are clearly and analytically presented in the following FIGURE 2 and TABLE 3 for the Foursquare dataset and in FIGURE 3 and TABLE 4 for the Gowalla dataset, respectively.

It is clearly shown in FIGURE 2 that DPGSR-PR is effective and consistently outperforms the first four base-lines for various values of $N$. Specifically, PMF is more efficient than Basic MF and Biased MF, and GeoCF can further improve upon PMF. DPGSR-PR outperforms GeoCF with significant margins for all three evaluation metrics. TABLE 3 shows the major improvements made by DPGSR-PR compared to GeoCF and CoRe with $N$=10. Furthermore, DPGSR-PR is better than CoRe in terms of $P@N$, $R@N$ and $NDCG@N$. We observe a 8.45% improvement for $P@N$, a 10.45% improvement for $R@N$ and a 6.49% improvement for $NDCG@N$. To sum up, experimental results show that once the deep potential geo-social relationships under a certain geographical region where the user can arrive are considered, the performance of DPGSR-PR gets improved.

F. ANALYSIS ON DIMENSION $K$ OF LATENT FACTORS

The impact of the dimension $K$ of latent factors is discussed in this subsection. We conduct experiments on $P@10$, $R@10$ and $NDCG@10$. The comparison trends are similar to those presented in FIGURE 2 and TABLE 3, DPGSR-PR still gets the best performance in terms of $P@N$, $R@N$ and $NDCG@N$. After all, our approach outperforms all other baselines showing the benefit of DPGSR-PR.
and \(NDCG@10\) on the Foursquare and Gowalla datasets. The dimension \(K\) of latent factors can affect the rank-approximation and factorization of similarity matrix. We vary \(K\) from \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}. Experimental results are presented in FIGURE 4.

The curves plotted in FIGURE 4 clearly show how the change of dimension influences the results on the Foursquare and Gowalla datasets. When \(K\) is less than 40, the values of all measures increase quickly with the rise of \(K\) and the growth rate is fast. When \(K\) is larger than 40, the ranking measures \(P@10\), \(R@10\) and \(NDCG@10\) start to converge and retain pretty stable values. \(K\) impacts the time complexity of our proposed model, as larger \(K\) involves more computation. We set the dimension of \(K = 40\) on the Foursquare and Gowalla datasets under simultaneously considering the performance and the computation cost. Following this mode, we manage to obtain acceptable results on three metrics.

G. ANALYSIS ON REGULARIZATION PARAMETERS \(\beta\) AND \(\gamma\)

The changes of parameters \(\beta\) and \(\gamma\) affect the accuracy of POI recommendation. A 10-fold cross validation is conducted to set the regularization parameters \(\beta\) and \(\gamma\), and analyze influence of \(\beta\) and \(\gamma\) on \(P@10\), \(R@10\) and \(NDCG@10\). Moreover, we select \(\beta\) and \(\gamma\) both from the set \{0, 0.01, 0.05, 0.1, 0.5, 1\} to obtain different combinations. Results on \(P@10\), \(R@10\) and \(NDCG@10\) are shown in FIGURE 5 on Foursquare dataset and FIGURE 6 on Gowalla dataset, respectively.

FIGURE 5 and FIGURE 6 show that the result distributions over \(\beta\) and \(\gamma\) are similar in the two datasets and also present that the parameters \(\beta\) and \(\gamma\) are very important, because \(\beta\) controls influence from the explicit-implicit geo-social relationships between users and \(\gamma\) determines the effects of the implicit common visited POI-based relationships. As can be seen from FIGURE 5 and FIGURE 6, when \(\beta\) and \(\gamma\) are both small, i.e. the left lower corner of each figure, experimental results do not improve significantly, which is due to the fact that the geo-social regularization and constraints on unobserved entities cannot provide enough side information, the low-rank matrix factorization plays the most important role. As \(\beta\) and \(\gamma\) continue to grow, the measured values increase and the geo-social regularization and constraints on unobserved entities play more and more important roles, which shows that the heterogeneous social regularization has positive effect on the recommendation performance. With the continuously rise of \(\beta\) and \(\gamma\), the geo-social regularization and constraints have a negative effect on the low-rank structure of user-POI matrix.

VII. CONCLUSION AND FUTURE WORK

This research introduces the deep potential geo-social relationship mining-based POI Recommendation model known as DPGSR-PR, which is based on kernel density estimation, two-hop random walk and Jaccard similarity coefficient. DPGSR-PR simultaneously take into account the explicit and implicit geo-social relationships and the implicit common check-in POI-based geo-social relationships between users (ICP-GSR). Experimental results clearly present that DPGSR-PR is better than other five established methods on two datasets, which is because DPGSR-PR is a unified method, and the involved two deep potential geo-social relationships are within a user reachable region.

There are some interesting researches to be continuously studied. First, in addition to consider the deep potential social relationship based on EIU-GSR and ICP-GSR under a certain geographical region, the user check-in behavior also change as time goes by. Hence, we will employ the time factor in POI Recommendation. Second, the reviews of some POIs from users greatly contribute to deeply mine the relationships between users. In addition, we are also attracted by the successive POI recommendation based on deep learning.
REFERENCES


FIGURE 5. Distributions of evaluation metrics on $\beta$ and $\gamma$ for the Foursquare dataset

FIGURE 6. Distributions of evaluation metrics on $\beta$ and $\gamma$ for the Gowalla dataset


XIANWEI LI has been with School of Information Engineering, Suzhou University, P.R.China, since July 2011. He is also a Ph.D candidate in Global Information and Telecommunication Institute at Waseda University, Tokyo, Japan. His research interests include resource management in cloud computing, computer networks and 5G communication networks.

GUOLONG CHEN is a professor in School of Information Engineering, Suzhou University, Anhui Province, P.R.China. He received master degree and Ph.D degree in 1993 and 1998, respectively, both in applied maths. His research interests include cloud computing, computer networks, and cognitive radio networks.

***