A Diverse and Personalized POI Recommendation Approach by Integrating Geo-Social Embedding Relations

XIANGFU MENG 1,2 AND JINFENG FANG 1

1School of Geomatics, Liaoning Technical University, Fuxin 123000, China
2School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China

Corresponding author: Jinfeng Fang (Infangziyi@163.com)

This work was supported by the National Science Foundation of China under Grant 61772249.

ABSTRACT User-POI rating matrix is one of the current research hotspot of POI recommendation algorithms, the goal of which is to obtain the POIs with the highest user satisfaction. However, most existing POI recommendation algorithms ignore the diversity of recommended POI list, and the POIs in the recommended list are usually similar to each other, which cannot effectively broaden the user’s perspectives. To solve this problem, a diverse and personalized recommendation method EBPRMF (Embedded Bayesian Personalized Ranking Matrix Factorization) that integrates the embedded features of the geographic-social relationships of POIs is proposed. First, by embedding and compressing the geographic and social relationships between POIs, a geographic-social relationship embedding model is constructed to evaluate the coupled correlation between POIs. The correlation between all POIs forms a POI correlation matrix. And then the spectral clustering method is leveraged to cluster the POIs according to the correlation matrix, and several different POI clusters can be obtained. Next, the Bayesian personalized ranking matrix factorization model is proposed to select POI that is close to the user’s preference from each cluster, and finally a recommendation list of POIs that is both diverse and personalized is obtained. Experimental results demonstrate that the geographic-social relationship embedding model can well represent the location and social relationship between POIs, and has a good clustering effect. The diversity of recommendation list is high, and the sorted list can be well satisfied with user preferences.

INDEX TERMS POI recommendation, geographic-social relationship embedding model, spectral clustering, diversity, Bayesian personalized ranking matrix factorization model.

I. INTRODUCTION

The development of geographic information systems and mobile networks has promoted the rapid development of Location-based Social Network (LBSN). With the increasing number of spatial web objects, also known as Points of Interest (POI), POI recommendation as one of the important services of LBSNs, has gradually become a hot topic in the current web query and recommendation research field. In 2016, Foursquare had more than 50 million active users and more than 8 billion POI check-ins information. The well-known American review site Yelp had approximately 21 million users and 102 million business reviews associated with geographic locations [1]. Sufficient data sources provide good opportunities for the research of POI recommendation. The most existing POI recommendation algorithms focus on fitting the user-POI satisfaction matrix. Based on the satisfaction matrix, the set of POIs closest to the current user’s preferences is selected and recommended to the user [2], [3], [4]. Although these methods ensure the accuracy of the recommendation results to a certain extent, they ignore the diversity of the recommendation results. The results obtained in this way do not consider the differences between POIs, which cannot satisfy the user’s diversity requirements. Therefore, a diversity and personalized POI recommendation system is becoming a research hotspot in the field of POI recommendation.

With the extensive application of personalization technology in Web database query, researchers have begun to study diverse Web data query and recommendation methods.
Deng et al. [5] make diversity recommendation through user preferences and dynamic interests. Jiang et al. [6] propose a search engine query recommendation method that comprehensively considers personalization and diversification. Cheng et al. [7] propose a recommendation method that takes into account product accuracy and diversity through supervised learning. Inspired by the personalization and diversity ideas in Web query and information retrieval, this paper integrates the two into the POI recommendation system, gets the user’s diverse tendencies by clustering user’s historical information, and then obtains the user’s personalized preferences through Bayesian personalized rating matrix factorization, finally achieves both diversity and personalized POI recommendation.

The contributions of this paper are as follows.

1) According to the geographic and social relationships of POIs, a geo-social relationship embedding model is proposed to make the geographic-social relationship correlation more reasonable.

2) Based on the relevance of the geo-social relationship of POI, a clustering method of POIs based on spectral clustering is proposed. The POIs in the same cluster are close to each other while they are far from each other in different clusters.

3) A personalized selection and ranking method is proposed by using the Bayesian personalized rating matrix factorization algorithm, which is used to provide users with a diverse and personalized POI recommendation list.

4) Preliminary experiments are conducted on large-scale real datasets to evaluate the effects and performance of the proposed algorithm. Experimental results demonstrated the effectiveness and superiority of the proposed algorithms.

The rest of this paper is organized as follows: Section 2 introduces related work; Section 3 presents the problem definition and solutions; Section 4 proposes the geographic-social relationship embedding model; Section 5 describes the spectral clustering of POIs; Section 6 proposes a personalized ranking method for POI recommendation; Section 7 is the experimental evaluation and the paper is concluded in Section 8.

II. RELATED WORK

In recent years, with the development of multimedia technology and the increasing number of mass POI data, POI recommendation algorithms have received widespread attention. POI recommendation is based on the evaluation of the relevance between POIs. So, we will introduce related work from two levels: POI relevance evaluation and POI recommendation.

A. POI RELEVANCE EVALUATION

The relevance evaluation of POI is the basis of POI recommendation. An effective relevance evaluation model plays a vital role in the recommendation process. The information associated to POI includes three categories, namely geographic information, text information, and social information. Geographic information refers to the location information of POI, usually expressed by latitude and longitude. Text information is a text description of POI attributes (or characteristics), such as POI name, POI category, etc. Social information mainly refers to the check-in information of POI, including check-in users, check-in time and user comments on POI, etc. There are mainly two types of POI correlation evaluation methods, the geographic-textual relationship evaluation model and the geographic-social relationship evaluation model. The geographic-textual relationship evaluation model (abbreviated as Geo-textual model) is a model based on the geographic information and text information associated to POI. The geographic relationship of POI is mainly computed by Euclidean distance or road network distance between POIs. The text relationships of POIs are mainly measured by using Cosine similarities between text descriptions between POIs. This model is widely used in POI query and recommendation [8], [9], [10], [11]. The geographic-social relationship evaluation model (abbreviated as Geo-social Model) is a very popular POI correlation evaluation model in recent years, and it is widely used in POI recommendation and spatial keyword query. The geographic relationship between POIs is similar to the Geo-textual model, and the social relationship is mainly measured by the relationships of users visiting the POIs. Several researchers have found that users with close relationships in social networks usually have similar check-in behaviors. Wang et al. [12] conduct experiments on two location-based social network datasets Brightkite and Gowalla. It found that in the two data sets the overlap ratios of the user’s first check-in location and their direct friend or indirect friend’s check-in location are 23% and 31%, respectively. Experimental results showed that user relationships have a great impact on their access to POIs. The method proposed in [13] injects the user’s Geo-social preferences into the POI recommendation process and leverages a three-level joint pairwise ranking model to reconstruct the POI recommendation model. Pan et al. [14] introduce a POI recommendation model by the deep potential Geo-social relations. With the help of two-hop random walk and Jaccard similarity coefficient, the explicit and invisible Geo-social relationship is considered at the same time to provide the recommendation of POI. Xu et al. [15] recommend POIs based on the user’s check-in data. They first introduce the 2-degree friend relationship into the collaborative filtering algorithm to build a social influence model, and then add POI location factors to build a location influence model. Next, they combine the two models together to form a Geo-social relationship model, and finally the kernel density estimation method is leveraged to improve the recommendation accuracy.

B. POI RECOMMENDATION

The current POI recommendation technologies can be divided into the following three categories, content-based recommendation, model-based collaborative filtering recommendation, and personalized recommendation.

The content-based recommendation algorithm mainly traverses user consumption data in the database, and makes
recommendations for current users based on similar users or POIs. Ye et al. [16] calculate the similarity between users, look for users with similar preferences to the current user, predict the current user’s satisfaction with POI based on similar users, and then recommend POIs for them. FUIT proposed by [17] combines the user preference models (UIM) and the real-time traffic model (RTTM) to realize a content-based POI recommendation under real-time road conditions. Huang et al. [18] exploit content-based recommendation algorithm to find a timing plan suitable for current traffic conditions. The traffic situation is regarded as the user, the timing plan is regarded as the POI, the traffic indicators such as delay time are regarded as the user’s rating of the POI, and similar traffic conditions are found through the content-based recommendation algorithm and the k nearest neighbor algorithm. The development of data mining technology and the progress of machine learning theory provide a good foundation for model-based collaborative filtering algorithm. The model-based collaborative filtering algorithm establishes an effective machine learning model by analyzing the user-item evaluation matrix, which alleviates the problem of data sparsity and scalability, and then obtains accurate recommendation results [19], [20]. Currently, implicit semantic model [21], [22], [23], clustering model [24]–[27] and Bayesian model [28]–[30] are commonly used in POI recommendation. Personalized recommendation is currently the mainstream development direction of POI recommendation. Recommendations can be made according to POI spatial-temporal attributes and user behavior patterns; it can also be directly learned from historical data such as the user’s check-in records, and combine the collaborative filtering and user behavior analysis to achieve personalized recommendations for POI. Han et al. [31] analyze users’ historical access data to POI, leverage location-based filtering algorithms to reduce noise and interference, and obtain recommendations that are close to user needs. ST-LDA provided by [32] learns region-dependent personal preferences based on the check-in information of POI in each region. Zhang et al. [33] consider that users’ geographic preferences usually change over time, and propose a personalized geographic impact modeling method PGIM for POI recommendation. Zhou et al. [34] exploit an adversarial POI recommendation model APOIR to explore the distribution of users’ potential preferences. The model consists of two parts: recommender R and discriminator D. The two components are jointly trained to integrate geographic and social relationships into the reward function, and optimize the recommender R in a reinforcement learning way to make the recommendation results closer to user preferences. Meng et al. [35] comprehensively consider the geographic and social relationships of POI, integrate both through linear weighting to construct an evaluation model for POI, and introduce the probability factor algorithm into the POI recommendation process to form a DPFM personalized POI recommendation model to realize the personalized recommendation of POI.

III. PROBLEM DEFINITIONS AND SOLUTION

This section first gives definitions related to POI recommendation, and then presents our solution framework.

A. DEFINITIONS

Definition 1 (POI Set): Let \( p_i = (\text{loc}, \text{doc}, \text{soc}) \) denote the POI \( p_i \), where \( \text{loc} \) is the geographic information of the POI, such as latitude and longitude; \( \text{doc} \) is the text information of the POI, such as name, category, etc.; \( \text{soc} \) is the social information of the POI, such as user check-in records, user comments, etc. All POIs constitute a collection of POI \( F = \{p_1, p_2, p_3, \ldots, p_n\} \).

Definition 2 (User Social Network): Let \( G = (U, E) \) represent the social network, where \( U \) represents the collection of all users, and the relationship between users constitutes the edge \( E \) of the social network. \( (u_i, u_j) \in E \) means that users \( u_i \) and \( u_j \) have a direct social relationship.

Definition 3 (User Check-in Record): Let \( CK = \{< u_i, p_k, t_r > | u_i \in U, p_k \in P \} \) represent the check-in records of all users in the user set \( U \) at all the POIs in the POI set \( P \). Among them, \( < u_i, p_k, t_r > \) represents that user \( u_i \) has visited POI \( p_k \) at time \( t_r \). Given a POI \( p_k \), the set of users who have visited the POI \( p_k \) can be represented as \( U_{pk} = \{u_i | < u_i, p_k, * > \in CK \} \), where \( * \) means any time.

B. PROBLEM SOLUTION

The solution proposed in this paper is shown in Figure 1. First, it builds a POI similarity evaluation model. The geographic information of POI and the social relationship information between POIs are embedded into a multi-dimensional space, in which a multi-dimensional vector is generated to represent the POIs. And then, the similarity between POIs can be obtained by calculating the Cosine value between the vectors. Consequently, the correlation matrix of all the pairs of POIs can be formed. Second, On the basis of the correlation matrix, spectral clustering is used to cluster POIs, the similar POIs are clustered into one category. Last, based on the similarity matrix obtained by embedding geographical information and social relationship information of POIs, the classic Bayesian ranking matrix factorization algorithm is adopted to select the optimal POI for each category of POIs, which forms an embedded Bayesian personalized ranking matrix factorization algorithm (EBPRMF). The optimal POIs selected from each cluster by EBPRMF are sorted in descending order according to user satisfaction, thereby obtaining both diversity and personalized recommendation results.

IV. GEO-SOCIAL RELATIONSHIP EMBEDDING MODEL

This section first describes the extraction of POI geographic and social relationship, and then proposes a novel integration method to construct the POI geo-social relationship model.

A. GEOGRAPHIC AND SOCIAL RELATIONSHIP EXTRACTION

This paper extracts the latitude and longitude of POI as the geographic information, which is used to describe the
geographic location of POI. The social relationship information is determined by the users who visit the POI, and inferred from the social relationship among the users who visit them. The social relationship of users is evaluated by the fact that users have visited POI together. The more similar the POIs visited by a pair of users are, the more similar the interests of the users are, and the closer the social relationship between the two POIs. The calculation of the social relationship of POI is divided into two steps. First, it calculates the similarity of social relations between users who have visited the POI set, and then calculates the similarity of social relations between POIs. The calculation process is as follows:

**Step 1:** Calculate the similarity of user social relationships. The calculation method for the similarity $S_{u_a,u_b}$ of the social relationship between a pair of users $u_a$ and $u_b$ is:

$$S_{u_a,u_b} = \frac{|P_{u_a} \cap P_{u_b}|}{|P_{u_a} \cup P_{u_b}|}$$  \hspace{1cm} (1)$$

where $P_{u_a}$ represents the set of POIs visited by user $u_a$, $P_{u_b}$ represents the set of POIs visited by user $u_b$. Table 1 shows the check-in records of 5 users to 6 POIs. Take Table 1 as an example to calculate the similarity of social relations between users.

From Table 1, we can see that the set of POIs visited by user $u_1$ is $u_1 = \{p_1, p_3, p_5\}$, the set of POIs visited by user $u_2$ is $u_2 = \{p_1, p_2, p_3, p_4\}$. Then we can calculate the similarity of the direct social relationship between two users in the user set by using Equation (1), and the results are shown in Table 2.

**Step 2:** Calculate the similarity of POI social relationships. The similarity of social relationships between two POIs is calculated on the basis of the social relations similarity of users who have visited the POIs. The similarity $ST_{p_ip_j}$ of the social relationship between two POIs $p_i$ and $p_j$ is:

$$ST_{p_ip_j} = \left\{ \begin{array}{ll} \frac{1}{|U_{p_i} \cup U_{p_j}|} \sum_{u_a \in U_{p_i}} \sum_{u_b \in U_{p_j}} S_{u_a,u_b}, & U_b \neq \emptyset \\ 0, & U_b = \emptyset \end{array} \right.$$  \hspace{1cm} (2)$$

where $U_{p_i}$ represents the set of users who have visited the POI $p_i$, and $U_{p_j}$ represents the set of users who have visited the POI $p_j$. $U_a = \{u_a | u_a \in \{U_{p_i} \cup U_{p_j}\}\}$

### Table 1. POIs check in records.

| POI | Latitude | Longitude | User |
|-----|----------|-----------|------|
| $p_1$ | 41.921738 | -87.712336 | $u_1,u_2,u_3$ |
| $p_2$ | 41.961353 | -87.755029 | $u_2,u_3,u_4$ |
| $p_3$ | 41.927651 | -87.705039 | $u_3,u_2,u_4,u_5$ |
| $p_4$ | 41.968939 | -87.728109 | $u_2,u_3$ |
| $p_5$ | 41.927944 | -87.705177 | $u_1,u_2,u_3$ |
| $p_6$ | 41.927906 | -87.705089 | $u_4,u_5$ |

### Table 2. User Relationship Similarity Matrix.

|      | $U_1$ | $U_2$ | $U_3$ | $U_4$ | $U_5$ |
|------|-------|-------|-------|-------|-------|
| $U_1$ | 1     | 2/5   | 2/5   | 1/5   | 1/2   |
| $U_2$ | 2/5   | 1     | 3/5   | 2/5   | 1/6   |
| $U_3$ | 2/5   | 3/5   | 1     | 2/5   | 1/6   |
| $U_4$ | 1     | 2/5   | 2/5   | 1     | 1/5   |
| $U_5$ | 1/2   | 1/6   | 1/6   | 1/5   | 1     |
TABLE 3. POIs Relationships Matrix.

|    | \(p_1\) | \(p_2\) | \(p_3\) | \(p_4\) | \(p_5\) | \(p_6\) |
|----|---------|---------|---------|---------|---------|---------|
| \(p_1\) | 1       | 0.4700  | 0.6233  | 0.4875  | 0.7208  | 0.5750  |
| \(p_2\) | 0.4700  | 1       | 0.8500  | 0.8000  | 0.4133  | 0.3750  |
| \(p_3\) | 0.6233  | 0.8500  | 1       | 0.7000  | 0.5467  | 0.4800  |
| \(p_4\) | 0.4875  | 0.8000  | 0.7000  | 1       | 0.5167  | 0.2833  |
| \(p_5\) | 0.7208  | 0.4133  | 0.5467  | 0.5167  | 1       | 0.4167  |
| \(p_6\) | 0.5750  | 0.3750  | 0.4800  | 0.2833  | 0.4167  | 1       |

B. GEO-SOCIAL RELATIONSHIP INTEGRATION

The linear weighting method is often used to integrate the Geo-social relationship of POI. First, it calculates the geographic distance and social relationship distance of two POIs, and then assigns a weight to each, and finally computes the similarity between POIs by summing these two kinds of distances [35], [36]. However, the proportional distribution of the two distances needs to be adjusted manually. For different situations, the proportions of the two distances should be different. To address this problem, we propose a new way to integrate the POI information. The specific integration procedure is divided into three steps. First, the POIs are represented by the form of vectors, and then the set of POIs can be transformed into a vector set. After this, the vectors of POIs are compressed to obtain an embedded vector representation. Finally, the Cosine similarity is used to calculate the value of the embedded vectors of two POIs.

Step 1 Vector Representation of POI: Each POI vector consists of two parts, that is, geographic information and social relationship information. The geographic information of POI is its latitude and longitude, in order to avoid the calculation result being too dependent on a certain attribute, we first use Equation (5) to simplify the latitude and longitude, where \(maxL\) is the value having the maximum absolute values of the latitude and longitude of POIs. For example, given a POI set as showed in Table 1, here the \(maxL\) is \(-87.755029\). Then, all the longitude and latitude values are scaled by dividing the value of \(maxL\).

\[
NL_{lat_i} = \frac{Lat_{p_i}}{maxL} \\
NL_{lon_i} = \frac{Lon_{p_i}}{maxL} \tag{5}
\]

The social relationship information of POI is its mapping in the POI set. Assuming that there are \(n\) POIs in the POI set, then each POI can be expressed as an \(n + 2\) dimensional vector. As shown in Figure 3(a), the first two bits of the POI vector \(p_i\) are its normalized latitude and longitude, and the last \(n\) dimensions are the mapping of the POI \(p_i\) on the \(p_j\) dimension \((p_1, \ldots, p_n)\), represented by the similarity between \(p_i\) and \(p_j\).

![FIGURE 2. User relationship graph.](image)

Take POI \(p_1\) and \(p_2\) as an example. From the user’s social relationship in Figure 2 and formula (4), when \(u_a = u_1\), \(U_b = \{u_2\}\); when \(u_a = u_2\), \(U_b = \{u_2\}\); when \(u_a = u_3\), \(U_b = \{u_2, u_3\}\); when \(u_a = u_4\), \(U_b = \{u_2\}\); when \(u_a = u_5\), \(U_b = \{u_3\}\). Through combining user’s social relationship similarities in Table 2, we get the similarity \(ST_{p_1p_2}\) of social relations between POI \(p_1\) and \(p_2\):

\[
ST_{p_1p_2} = 1/5^4(2/5 + 1 + (3/5 + 1/6)/2 + 2/5 + 1/6) = 0.4700
\]

In the same way, the similarity of the social relationship between two POIs is shown in Table 3.
Step 2 Embedded Vector Representation of POI: for a large scale of dataset, the cardinality of POI (represented by \( n \)) would be large, and the dimension represented by the POI vector would also be large. To reduce the calculation, the POI vector representation will be multiplied by a matrix \( T \) with a dimension of \((n+2)^m\). Thereby the original \( n+m \) dimension POI vector is represented as an \( m \)-dimensional compressed vector, where \( m \) is the ideal vector length, as shown in Figure 3 (b). To make the embedded POI vector closer to the original POI vector, multiple iterations are used to select the optimal value to optimize the matrix \( T \). We first initialize the matrix \( T \) randomly. To ensure the uniformity of data, the value of the matrix \( T \) should satisfy the symmetry centered at 0 as much as possible. And then, we randomly pick up some POIs and calculate the sum of the mean square error between the original POI vector cosine similarity and the compressed POI vector cosine similarity, which is called the compression cost, as in Equation (6). The embedded POI vector should be as close to the original POI vector as possible, that is, the smaller the compression cost the better. After some iterations of optimization, an ideal compressed matrix \( T \) can be obtained. By multiplying the POI vector by the compressed matrix \( T \), the final embedded vector representation of all POIs can be obtained.

\[
\text{Cost} = \frac{1}{n} \sum_{i,j \in LP} \left( SM_{p_ip_j} - \hat{SM}_{p_ip_j} \right)^2
\]

where \( LP \) is the partial POI set selected randomly; \( SM_{p_ip_j} \) is the cosine similarity between the original vectors of POI \( p_i \) and \( p_j \); \( \hat{SM}_{p_ip_j} \) is the cosine similarity between the embedded vectors of POI \( p_i \) and \( p_j \); \( \hat{n} \) is the number of POIs in \( LP \).

Assuming that the ideal vector dimension \( m \) in the example in Section 4.1 is 3, the embedded vector representation of each POI can be obtained as shown in Table 4.

### Table 4. Embedded vector representation of POIs.

| POI | Embedded vector representation |
|-----|--------------------------------|
| \( p_1 \) | (-0.26159523, -0.077249, 0.88791497) |
| \( p_2 \) | (0.1361287, -0.31737358, 0.91441508) |
| \( p_3 \) | (-0.01330655, -0.24321413, 0.97160729) |
| \( p_4 \) | (0.10882983, -0.27645919, 0.8284637) |
| \( p_5 \) | (-0.12889149, 0.015906, 0.79105736) |
| \( p_6 \) | (-0.37831701, -0.35799137, 0.85381965) |

Step 3 Similarity Evaluation of POI: We use the cosine similarity to measure the POI Geo-social relationships. The overall similarity between POI \( p_i \) and \( p_j \) is calculated as follows:

\[
SM_{p_ip_j} = \frac{\sum_{r=1}^{m} (p_i^r) \cdot (p_j^r)}{\sqrt{\sum_{r=1}^{m} (p_i^r)^2} \cdot \sqrt{\sum_{r=1}^{m} (p_j^r)^2}}
\]

where \( m \) is the ideal dimension, \( p_i^r \) is the value of the embedded POI \( p_i \) in the \( r \)-th dimension. On the basis of Step 1, a correlation matrix \( SM \) of \( n \times n \) can be obtained by using Equation (7). The elements in the matrix \( SM_{p_ip_j} \) are the Geo-social relationship correlation between POI \( p_i \) and \( p_j \). The similarity evaluation matrix of POIs in the example in Section 4.1 is shown in Figure 4 (Figure 4(a) is the similarity evaluation matrix before embedding, and Figure 4(b) is the similarity evaluation matrix after embedding).

![FIGURE 4. POIs similarity evaluation matrix.](image)

V. SPECTRAL CLUSTERING

The goal of clustering is to aggregate similar observations into clusters. The elements in the same clusters are as similar as possible and the elements between clusters are as different as possible. Spectral clustering is an effective method commonly used in cluster analysis. Compared with other clustering algorithms, it has stronger adaptability and more accurate clustering effect. Spectral clustering evolved from graph theory. The main idea is to treat data as points in the graph, the relationships between data points as edges in the graph, and the closeness of data points is represented by the weight of the edges in the graph. The closer the data point relationship, the greater the weight of the edge between data points, and vice versa, the smaller. In this way, data points and graph are connected, and the clustering of data points can be transformed into the cutting of the graph. The goal of spectral clustering is to cut the data graph so that the weight of the inner edge of the same subgraph is as high as possible, and the weight of the edge between different subgraphs is as low as possible [37]. Spectral clustering is simple in operation and convenient in calculation, suitable for the division of high-dimensional spatial data, and only need the information of vertices and edges (i.e. the relationships of data points). The similarity evaluation matrix of POIs in this paper meets the characteristics of spectral clustering, and thus we take advantage of spectral clustering to partition the POI set.

In this paper, a two-way spectral clustering method is used. First, the relationship network graph of POI is decomposed into two optimal subgraphs \( G_1 \) and \( G_2 \), and the corresponding
cluster labels are $g_1$ and $g_2$, respectively. And then, we mark $y$ to record the division result of the relationship network graph, then $y = [y_1, y_2, \ldots, y_n]$, where $n$ is the total number of POIs. If the vertex $i$ in the division result belongs to $G_1$, then $y_i = g_1$, if the vertex $i$ belongs to $G_2$, then $y_i = g_2$. Therefore, the division scheme can be represented as $y = \{g_1, g_2, \ldots, g_n\}$, and the value of $ord$ is 1 or 2. We take the weights sum of the edges truncated when dividing the subgraph as the loss function. The loss function is as follows:

$$Cut(G_1, G_2) = \sum_{i\in G_1, j\in G_2} M_{p,p_j} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} M_{p,p_j} (y_i - y_j)^2}{2(g_1 - g_2)^2} \tag{8}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} M_{p,p_j} (y_i - y_j)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} M_{p,p_j} (y_i^2 - 2y_i y_j + y_j^2) = -\sum_{i=1}^{n} \sum_{j=1}^{n} 2M_{p,p_j} y_i y_j + \sum_{i=1}^{n} \sum_{j=1}^{n} M_{p,p_j} (y_j^2 + y_j^2) = -\sum_{i=1}^{n} \sum_{j=1}^{n} 2M_{p,p_j} y_i y_j + 2\sum_{i=1}^{n} y_i^2 \sum_{j=1}^{n} M_{p,p_j} = 2y^T (D - M)y \tag{9}$$

where $M$ is the similarity matrix, and the diagonal element is 0 (because the POI can be not connected to itself). The remaining elements are the same as the POI similarity evaluation matrix $SM$, $D$ is a diagonal matrix, and $D_j = \sum_{p=1}^{n} M_{p,p_j}$. By simultaneously considering the Equation (8) and (9):

$$Cut(G_1, G_2) = \frac{y^T Ly}{(g_1 - g_2)^2} \tag{10}$$

Among them, $L=D-M$ is called the Laplace matrix. It can be seen from Equation (10) that when the minimum value of $\lambda_{min}$ is obtained, the division loss function of the weighted undirected graph obtains the minimum value.

The POI set is divided by spectral clustering according to the minimum cut set criterion [38]. Let cluster label $g_1 = g$, $g_2 = -g$, then $y^T y = \sum_{i=1}^{n} y_i^2 = ng^2$, the cutting loss function of POI relationship graph can be formulated as (11):

$$Cut(G_1, G_2) = \frac{y^T Ly}{4g^2} = \frac{y^T Ly}{4y^T y} \tag{11}$$

The transformed loss function satisfies the Rayleigh quotient definition [39]. According to the property of Rayleigh quotient, when $y$ is the eigenvector corresponding to the smallest eigenvalue, the second smallest eigenvalue, the largest eigenvalue, the second largest eigenvalue, the minimum value, second smallest value, maximum value of $Cut(G_1, G_2)$ and then get the best division scheme that satisfies the POI network graph can be obtained. According to the corresponding relationship between the obtained eigenvalue and the eigenvector, each vertex can be represented by the eigenvector of the Laplacian matrix $L$, and the eigenvector element combination corresponding to the first $r$ smallest eigenvalues is used to represent each vertex coordinate. For example, if we want to divide the network graph into $k$ subgraphs, we can take the eigenvectors corresponding to the first $r$ smallest eigenvalues to form an $n \times n$ matrix $R$, and the $i$-th row vector represents vertex $i$, then perform $k$-means clustering on the vertices, finally get $k$ subgraphs (that is, $k$ POI clusters).

Take the 6 POIs in Section 4.1 as an example to illustrate the process of POI clustering. Figure 5 is their relationship network. The vertices $p_1$-$p_6$ represent POIs $p_1$-$p_6$ respectively, and the connections between the vertices represent the correlation between the POIs, POIs with weaker correlation are indicated by dotted lines for easy viewing.

**FIGURE 5. POIs relationship network diagram.**

First, the algorithm decomposes the POI relationship network graph in Figure 5 into two optimal subgraphs $G_1$ and $G_2$. Assume that $p_1$, $p_5$, $p_6$ in the division result are the first type, namely $\{p_1, p_5, p_6\} \in G_1$; and $p_2$, $p_3$, $p_4$ are the second type, namely $\{p_2, p_3, p_4\} \in G_2$, then $y = [g, -g, -g, -g, g, g]$. The similarity matrix $M$, diagonal matrix $D$, and Laplacian matrix $L$ are shown in Figure 6. The eigenvalues and eigenvectors corresponding to the Laplacian matrix $L$ are shown in Table 5.

**FIGURE 6. POIs clustering matrix.**

We can observe from Table 5 that if only the feature vector corresponding to the smallest eigenvalue (i.e. 2.17e-16) is used for classification, the values of all 6 POIs are 0.4082, which obviously cannot be divided. Therefore, this example takes the eigenvectors corresponding to the first two smallest
TABLE 5. Properties of the test dataset.

| Eigenvalue | Eigenvector  |
|------------|-------------|
| 2.17e-16   | 0.4082      |
| 5.3460     | 0.4082      |
| 5.4743     | -0.4755     |
| 5.6374     | -0.0686     |
| 5.6445     | 0.7617      |
| 5.7665     | -0.1769     |

When \( k \) clusters are needed, it is only necessary to cluster the vector set of the above vertex mapping through the \( k \)-means algorithm. After clustering the POI set according to the above process by spectral clustering, the POIs closest to the user’s preference are selected from each cluster to form a recommendation list, and the POIs contained in the recommendation list are diverse.

VI. PERSONALIZED RANKING OF POIs

Personalized POI recommendation aims to accurately arrange the POIs in the recommendation list according to the user preferences [40]. Matrix factorization (MF) is the most widely used method in the recommendation system, the main idea of which is to decompose a rating matrix into the product of two matrices of items and users. The core assumption of it is to use implicit semantics (hidden variables) to express users and items, and their product relationship becomes the original element. These hidden variables represent the characteristics shared by the user and the item, which appear as attribute characteristics on the item and preference characteristics on the user. So the recommendation problem can be transformed into how to obtain two optimal small matrices. In POI recommendation, we are more concerned about which POI has higher priority for users (i.e. the ranking is higher). Therefore, we need a ranking algorithm which can sort all POIs corresponding to each user according to their preferences. Bayesian Personalized Ranking (BPR) is such a sorting algorithm suitable for our case. BPR is an optimization framework that uses stochastic gradient descent to achieve pairwise sorting [41]. BPR algorithm belongs to pairwise approach, and is optimized for each user’s product preferences. Pairwise sorting is better than single-sample optimization. Taking into account the dominance of MF in the recommendation system, we take advantage of MF as the recommendation model and leverage BPR to optimize it, and make assumptions about the probability distribution of model variables to form a Bayesian personalized ranking matrix factorization (BPRMF).

MF splits the high dimensional rating matrix into two low dimensional implicit matrices and performs product operations on them, optimizes the mean square error between the original matrix and the product matrix to learn two implicit matrices, and finally completes the recommendation to users. Take Equation (12) as the objective function, we find out the potential user and POI feature matrix by solving its optimal value, which embed users and POIs in the \( k \)-dimensional implicit feature vector space to approximate the regularized real matrix.

\[
\min_{W,H} \sum_{(u,p) \in D} (c_{u,p} - \hat{c}_{u,p})^2 + \lambda (\|W\|_2^2 + \|H\|_2^2) \quad (12)
\]

where \( W \) and \( H \) are the \( k \)-dimensional implicit feature matrix after the users and POIs are embedded respectively; \( \lambda \) is a hyper-parameter to avoid overfitting, which can be determined by means of cross-validation; \( c_{u,p} \) is the real score of user \( u \) for POI \( p \) (here refers to the frequency of visits); \( \hat{c}_{u,p} = w_u^T h_p^T \) represents the predicted visits of user \( u \) to POI \( p \), which can be modeled in a variety of ways, the modeling method used in this paper is shown in Equation (13).

\[
\hat{c}_{u,p} = \alpha + \beta_u + \beta_i + W_u^T H_i \quad (13)
\]

where \( \alpha \) is the global offset; \( \beta_u \) and \( \beta_i \) are the offsets of the user \( u \) and the POI \( p \) respectively.

We next apply BPR algorithm for optimization. We assume the training set \( D \) as a triple form of \( (u, p_1, p_2) \), where \( u \) represents the user, \( p_1 \) represents the POI that received positive feedback, \( p_2 \) represents the POI that is not feedback. Then the training set \( D \) can be calculated by using Equation (14):

\[
D = \{(u, p_1, p_2) | u \in U \land p_1 \in P_u^+ \land p_2 \in P \land P_u^+ \} \quad (14)
\]

Let \( \Theta \) represents the parameters of the model, \( \hat{c}_{u,p_1,p_2}(\Theta) \) describes the relationship between the triples \( (u, p_1, p_2) \), assuming that the user’s preferences are independent of each other, Equation (15) can be obtained as follows:

\[
\prod_{u \in U} P (\geq u | \Theta) = \prod_{(u,p_1,p_2) \in D} P (i > u | j | \Theta) = \prod_{(u,p_1,p_2) \in D} \sigma (\hat{c}_{u,p_1,p_2}(\Theta)) \quad (15)
\]

where \( \sigma \) is the sigmoid function that can be calculated by \( \sigma(x) = 1/(1 + \exp(-x)) \), \( \hat{c}_{u,p_1} = \hat{c}_{u,p_1} - \hat{c}_{u,p_2} \) can get
from Equation (13); \( p \) represents the preference of user \( u \), for example, \( i > u j \) indicates that compared with POI \( p_j \) the user \( u \) is more inclined to POI \( p_i \).

Taking the example in Section 4.1 as an example, the check-in records of 5 users at 6 POIs can be sorted as shown in Figure 7(a). According to BPR, taking user \( u_i \) as an example, his/her preference for POI \( p_1 \) to \( p_6 \) can be obtained as shown in Figure 7(b).

For a user \( u \), we expect that the feedback POI \( p_i \) is more likely than the non-feedback POI \( p_j \) and the greater the difference between the POI \( p_i \) and \( p_j \) the better. By considering the Equation (15), we can derive the Equation (16) as follows:

\[
\ln P(\Theta | >_u) \propto \ln P(>_u | \Theta)p(\Theta) = \ln \prod_{(u,p_i,p_j)\in D_s} \sigma \left( \frac{\hat{c}_{u,p_i,p_j}}{\Theta} \right) + \text{ln}p(\Theta) \tag{16}
\]

When the prior distribution of \( \Theta \) is a normal distribution, in fact, a regular term is added to the loss function. We leverage pairwise sorting to optimize the model with Equation (17) as the optimization criterion,

\[
L(\Theta) = \sum_{(u,p_i,p_j)\in D_s} \left[ \ln \sigma \left( \frac{\hat{c}_{u,p_i,p_j}}{\Theta} \right) - \lambda \| \Theta \|^2 \right] \tag{17}
\]

After this, we randomly extract triples \((u, p_i, p_j)\) from the training set \( D_s \), and use the following stochastic gradient descent method to update related parameters, where \( \eta \) is the learning rate.

\[
\Theta \leftarrow \Theta + \eta \left[ \sigma \left( \frac{\hat{c}_{u,p_i,p_j}}{\Theta} \right) \frac{\partial \hat{c}_{u,p_i,p_j}}{\partial \Theta} - \lambda \Theta \right] \tag{18}
\]

\[
\frac{\partial}{\partial H_i} \hat{c}_{u,p_i,p_j} = H_i - H_j, \quad \frac{\partial}{\partial p_i} \hat{c}_{u,p_i,p_j} = 1, \quad \frac{\partial}{\partial p_j} \hat{c}_{u,p_i,p_j} = -1
\]

\[
\frac{\partial}{\partial W_u} \hat{c}_{u,p_i,p_j} = W_u, \quad \frac{\partial}{\partial H_j} \hat{c}_{u,p_i,p_j} = -W_u \tag{19}
\]

The process is as follows:

In conjunction with Section 5, the optimal POI from each cluster can be picked according to this BPRMF model, and then the selected POIs in each cluster are collected and arranged in descending order to form the final recommendation list.

**Algorithm 1 BPRMF Model**

**Input:** Training set \( D_s \), learning rate \( \eta \), regularization parameter \( \lambda \), decomposition dimension \( k \)

**Output:** Recommendation list \( L \)

1. Initialize parameters \( \Theta \).
2. Extract triples \((u, p_i, p_j)\) from the training set \( D_s \).
3. Calculate the predicted visits \( \hat{c}_{u,p_i,p_j} \) and \( \hat{c}_{u,p_j,p_i} \) of the user \( u \) to POI \( p_i \) and \( p_j \) according to Equation (13), and obtain \( \hat{c}_{u,p_i,p_j} \).
4. Optimize model parameters \( \Theta \) by Equation (18), until the algorithm converges.
5. Calculate the user’s predicted visits to POI based on the parameters \( \Theta \), sort them in descending order, select the first \( l \) POIs (\( l \) is the number of POIs required) to generate a recommendation list \( L \).

**VII. EXPERIMENTS**

This section introduces the experimental dataset and setting, and then reports the experimental results.

**A. DATASETS**

The experimental dataset is the user check-in data of Gowalla from Stanford Large Network Dataset Collection between February 2009 and October 2010, including the spatio-temporal data of the check-in records of the users to POI and the user’s social relationship network graph. User’s check-in record is mainly composed of user id, check-in time, POI id, latitude and longitude of the POI, including 6,442,890 records from February 2009 to October 2010; User social relationship network graph is composed of 196,591 vertices and 950,327 edges, where the vertices represent users and the edges represent the social relationship between two users. In order to alleviate data sparsity, this article intercepts two parts of the dataset located in Chicago, the United States and Tokyo, Japan. In addition, the data is simply preprocessed by discarding users who have checked in less than 5 times in the dataset and POIs that have been visited less than 5 times. The preprocessed experimental dataset is shown in Table 7.

**B. EXPERIMENTAL SETTING**

The embedded feature POI recommendation algorithm evaluation criteria: the experiment utilizes two indicators of intra-cluster variation and inter-cluster variation to evaluate the performance of POI similarity evaluation. While the diversity and accuracy is leveraged to evaluate the effectiveness of POI recommendation.
The intra-cluster variation represents the divergence between POIs in the same cluster, which is calculated by the average distance of all POIs in the same cluster. The smaller the intra-cluster variation the better. The inter-cluster variation represents the divergence between POIs in different clusters. The inter-cluster variation of any two clusters is calculated by the distance between the cluster centers of the two clusters, the greater the inter-cluster variation the better.

We calculate the variation by Equation (20):

\[
E_r = \frac{CN^2}{\sum_{i=1}^{CN} \sum_{j=1}^{CN} \sum_{r=1}^{R} (p_{i}^{(r)} - p_{j}^{(r)})^2}
\]

(20)

where \( p_{i} \) and \( p_{j} \) are POIs; \( R \) represents the dimension of POI, \( p_{i}^{(r)} \) and \( p_{j}^{(r)} \) represents the value of POI \( p_{i} \) and \( p_{j} \) in the r-th dimension; When the POI \( p_{i} \) and \( p_{j} \) are in the same cluster, \( CN \) refers to the total number of POIs in the cluster where \( p_{i} \) and \( p_{j} \) are located, the \( E_r \) obtained at this case is the intra-cluster variation; when the POI \( p_{i} \) and \( p_{j} \) take different cluster centers respectively, then \( CN \) refers to the total number of the clusters, the \( E_r \) obtained is the inter-cluster variation.

Diversity represents the richness of the recommendation list, the higher the diversity the better. The measurement of diversity is defined as follows:

\[
Div_{L_{rec}(u_a)} = \frac{\sum_{p_{i} \in L_{rec}(u_a)} \sum_{p_{j} \in L_{rec}(u_a)} D_{gs}(p_{i}, p_{j})}{|L_{rec}|^2}
\]

(21)

\[
Div = \frac{\sum_{u_a \in U} Div_{L_{rec}(u_a)}}{|U|}
\]

(22)

where \( L_{rec}(u_a) \) indicates the POI recommendation list obtained for user \( u_a \); \( U \) is the user set; \( D_{gs}(p_{i}, p_{j}) = 1-\text{SM}_{p_ipj} \) represents the Geo-social distance between \( p_{i} \) and \( p_{j} \), and \( \text{SM}_{p_ipj} \) represents the Geo-social similarity between \( p_{i} \) and \( p_{j} \), which can get from Equation (7). The larger the value of \( Div_{L_{rec}(u_a)} \), the higher the diversity of the recommendation list \( L_{rec}(u_a) \); the larger the value of \( Div \), the higher the overall diversity of recommendation results.

Accuracy represents the correctness of the recommendation list, which is measured by the overall precision rate, the overall recall rate and the \( F1 \) value composed of the first two. The \( F1 \) value is a comprehensive evaluation index for the precision and recall rate of the recommended results, which can more comprehensively reflect the accuracy of the recommended list. The higher the precision rate, recall rate and \( F1 \) value, the better the accuracy of the recommended results. The calculation methods of accuracy are shown as follows:

\[
\text{precision} = \frac{\sum_{u_a \in U} L_{test}(u_a) \cap L_{rec}(u_a)}{|L_{rec}(u_a)|}
\]

(23)

\[
\text{recall} = \frac{\sum_{u_a \in U} L_{test}(u_a) \cap L_{rec}(u_a)}{|U|}
\]

(24)

\[
F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

(25)

where \( L_{test}(u_a) \) represents test POI list, which is composed of the top \( l \) POIs selected in the test set with the most visits by the user \( u_a \), and \( l = L_{test}(u_a) = L_{rec}(u_a) \).

To better demonstrate the effect of the recommendation algorithm proposed in this paper, the following algorithms are respectively compared with our method in the experiments.

**EBPRMF:** An improved algorithm of BPRMF proposed in this paper. It modified the POI similarity measurement, increased the accuracy of the BPRMF recommendation algorithm, and introduced spectral clustering to the traditional BPRMF recommendation algorithm to achieve recommendation diversity.

**BPRMF:** The algorithm is based on the MF algorithm and leverages BPR network for optimization. POI \( p_{i} \) (the POI receives feedback) and \( p_{j} \) (the POI receives non-feedback) are considered in pairs, then maximize the difference between the two to achieve personalized selection and recommendation of POI.

**DPFM:** The algorithm is based on the probability factor model (PFM), and realizes the diversity and personalized POI recommendation through linear weighting of Geo-social relations.

**DSVD:** The algorithm is based on singular value decomposition (SVD), and realizes the diversity and personalized POI recommendation through linear weighting of Geo-social relations.

**DNMF:** The algorithm is based on non-negative matrix factorization (NMF), and realizes the diversity and personalized POI recommendation through linear weighting of Geo-social relations.

\section*{C. ANALYSIS OF RESULTS}

\subsection*{1) SETTING OF PARAMETERS}

This paper first embeds and compresses the Geo-social information of all POIs, constructs a POI similarity evaluation matrix, and then divides the POI set by spectral clustering based on the evaluation matrix, finally achieves POI recommendation through BPRMF. The process involves three parameters: the embedding dimension \( m \) in the construction of POI similarity evaluation matrix in section 4.2; the number of minimum eigenvalues \( r \) of matrix \( L \) in spectral clustering in section 5; and the implicit matrix dimension \( k \) used in Bayesian personalized matrix decomposition in section 6. The complexity of EBPRMF should not be too high or too low, otherwise it will cause overfitting or reduce the likelihood of observations. We set the number of recommended POI to 10, so the number of subgraphs in spectral clustering is 10. In addition, the setting of other parameters in this paper refers to the [42], i.e., \( \eta = 0.005 \), \( \lambda = 0.02 \). To ensure the validity of the recommendation results, the final result is the average of 10 experiments results. The selection of the parameters \( m \), \( r \), and \( k \) will be discussed below. Figure 8 shows the compression cost corresponding to the embedding dimension \( m \).
Figure 8 shows the compression cost corresponding to the embedding dimension \( m \) when training set accounts for 50%, 60%, 70%, 80% and 90% of the two city (Chicago and Tokyo) datasets respectively. The compression cost refers to the divergence between the unembedded similarity matrix and the embedded similarity matrix of POI (see Equation (6) for the details). The smaller the value, the closer the embedded feature is to the original feature, and the more similar the information is to the real POI information. Figure 8(a) is the result obtained on the Chicago dataset, and Figure 8(b) is the result obtained on the Tokyo dataset. Considering that too small embedding dimensions will cause greater deviations, and too much dimensions will increase the computational burden, so the range of embedding dimensions set in this paper is between 100 and 500. From the experimental results, we find that the embedded dimension is not the bigger the better. In Chicago, 80% of the cases have the lowest compression cost when the embedding dimension \( m \) is set to 300; in Tokyo, 60% of the cases have the lowest compression cost when the embedding dimension \( m \) is set to 300. On the whole, the effect is best when the embedding dimension \( m \) is set to 300, so the embedding dimension \( m \) is set to 300.

In this experiment, we fix the embedding dimension \( m = 300 \), and test the impact of different number \( r \) of minimum eigenvalues on recommendation performance. The range of \( r \) is set between 1 and 10. Figure 9 shows the variation results of the average diversity and accuracy of POI recommendation with the parameter \( r \).

Figure 9, the accuracy of POI recommendation is reflected by F1 value, which is a comprehensive index of precision and recall, because F1 value can show the POI recommendation accuracy from the two perspectives of precision and recall. We observe that, for Chicago dataset, the diversity of POI recommendation achieves the highest accuracy when the number of minimum eigenvalues \( r \) is 9. In contrast, the recommendation results of Tokyo is very low in diversity. In addition, when the number of minimum eigenvalues \( r \) is 6 or 10, although the diversities of the recommended results on both Chicago and Tokyo are high, the F1 value obtained by the two cities are not high enough. It shows that although the diversity of the recommended results is satisfied, the accuracy cannot be guaranteed. When the number of minimum eigenvalues \( r \) is 2, the diversity and F1 value in the recommended results are both the highest in Tokyo; the F1 value is the highest and diversity is not low in Chicago. On the whole, it is more appropriate for \( r \) to be 2, so this paper sets the number of minimum eigenvalues \( r = 2 \).

Next, We set embedding dimension \( m = 300 \), the number of eigenvalues \( r = 2 \), and explore the influence of different implicit matrix dimensions (i.e. the number of hidden factors) \( k \) on the accuracy of POI recommendation. Adopts the average value of 5 situations with different proportions of the training set (50%, 60%, 70%, 80% and 90%) to avoid the contingency of the results. The values of hidden factors number \( k \) are set to 5, 10, 15, 20, 25, and the experiment results are shown in Figure 10.
FIGURE 10. The impact of k on the precision and recall of POI recommendation.

Figure 10(a), 10(b) show the effect of different k values on the precision and recall of recommended results respectively. It can be seen from Figure 10 that when 10 or 15 hidden factors are set, the precision and recall rate of POI recommendation are better, and the effect is best when k is 10, so the number of hidden factors k in this paper is set to 10.

Comparison of the Similarity Evaluation Effect between Embedding and Linear Combination Method

To verify the efficiency of the similarity evaluation of embedding method proposed in this paper, 50%, 60%, 70%, 80% and 90% of the dataset are selected as the training set, and the remaining 50%, 40%, 30%, 20%, 10% are used as the test set, and the clustering results obtained by EBPRMF, BPRMF, DPFM, EDPFM, DPFM, EDSP, DSVD, EDNMF and DNMF algorithms are evaluated on the four indicators of diversity, precision rate, recall rate and F1 value. These two variations of each algorithm are shown in Table 8 and Table 9, respectively (The former of each column is the experimental result of the Chicago dataset, and the latter is the experimental result of the Tokyo dataset). Note that the intra-cluster variation obtained from Chicago in Table 8 is too small to clearly observe the difference, so this set of data is multiplied by $10^{-2}$.

From the results of the Chicago data set in Table 8, it is obvious that the intra-cluster variation of the clustering results obtained by embedding method to construct the Geo-social relationship model is much smaller than that of the traditional linear combination method. The divergence in Tokyo is not as obvious as in Chicago, but the intra-cluster variation obtained by embedding method is still smaller than traditional method; In Table 9, whether in Chicago or Tokyo, the clustering results obtained by embedding method have greater inter-cluster variation than the traditional method, indicating that the embedding method is effective.

2) COMPARISON OF THE RECOMMENDATION EFFECT BETWEEN EBPRMF AND OTHER ALGORITHMS

To verify the effectiveness of EBPRMF in this paper, 50%, 60%, 70%, 80% and 90% of the dataset are selected as the training set, and the remaining 50%, 40%, 30%, 20%, 10% are used as the test set, and the effectiveness obtained by the BPRMF, DPFM, DSVD and DNMF algorithms are evaluated on the four indicators of diversity, precision rate, recall rate and F1 value. The diversity, precision rate and recall rate obtained by the five algorithms are shown in Table 10, Table 11 and Table 12, respectively (The former of each column is the experimental result of the Chicago dataset, and the latter is the experimental result of the Tokyo dataset). Finally, the F1 value curve is drawn for the recommended results obtained by the five algorithms, as shown in Figure 11.

Obviously, regardless of Chicago or Tokyo, the diversity of EBPRMF recommendation results are the highest in different training sets. Compared with other algorithms, the average diversity of the datasets in two cities has increased by 11.70%, 16.03%, 35.03%, and 31.40% respectively. Results show that the proposed algorithm is effective in improving the diversity of POI recommendation.

Table 11 records the precision rate comparison results of five recommendation algorithms such as EBPRMF in different city datasets, different proportions of training set and test set. The overall precision rate of the experimental results is low because the POIs users have visited are limited, and the dataset is relatively sparse. Besides, it can be found that among several algorithms, regardless of the proportion of the training set, the precision rate of EBPRMF is the highest. Its average precision rate in the two cities is 32.77%, 30.11%, 50.74% and 43.36% higher than the other four algorithms respectively, which shows that EBPRMF is also very valuable in ensuring the precision of the recommended results.

Finally, the F1 value curve is drawn for the recommended results obtained by the five algorithms, as shown in Figure 11.
and 37.48% higher than the others respectively; in the Tokyo dataset, the average recall rate of EBPRMF is slightly lower than that of the DPFM algorithm, only lower than 1.54%, but it is 67.14%, 15.86% and 13.14% higher than the remaining other three algorithms respectively; and in terms of the total average recall rate of different training sets in the two cities, the total average recall rate of EBPRMF is 14.65% higher than that of the DPFM algorithm. Therefore, the recall rate of EBPRMF is generally the best among several algorithms.

F1 value is a comprehensive evaluation index for precision and recall, which is used to reflect the overall effect of the algorithm. Figure 11(a) is the F1 values obtained by five algorithms such as EBPRMF on the Chicago dataset according to different proportions of the training set. Obviously, for the Chicago dataset, the DSVD has the worst effect and EBPRMF has the best effect. In addition, when the proportion of the training set is 50%, 60% and 90%, the effect of DPFM and DNMF are relatively similar; Figure 11(b) shows the F1 values obtained by five algorithms on the Tokyo dataset according to different proportions of the training set. We can see that when the proportion of the training set is 50%, EBPRMF and DPFM are almost the same, but when the scale of training set changes, there appears a gap between the two. In Tokyo dataset, BPRMF has the worst effect, but the best effect is still EBPRMF proposed in this paper.

In order to compare the overall pros and cons of each algorithm more intuitively, the average F1 values of the five algorithms on Chicago and Tokyo datasets are plotted as Figure 12. It is not difficult to find from Figure 12 that the effect of the algorithm in Tokyo is slightly better than that in Chicago. The reason is that the Tokyo dataset is more complete than

| Training set | EBPRMF | BPRMF | DPFM | DSVD | DNMF |
|--------------|--------|-------|------|------|------|
| 50%          | 0.7086/0.6248/0.5465/0.4780/0.4920/ | 0.7912/0.6946/0.6983/0.5735/0.5915/ | 0.6824/0.5874/0.5499/0.4715/0.4912/ | 0.8183/0.6663/0.6899/0.5722/0.5793/ | 0.6472/0.5914/0.5459/0.4879/0.4887/ | 0.7797/0.6862/0.6791/0.5685/0.5838/ | 0.6659/0.6073/0.4256/0.3572/0.3937/ | 0.7140/0.7116/0.8204/0.7294/0.7458/ | 0.6473/0.5537/0.5480/0.4988/0.5082/ | 0.7266/0.7063/0.6875/0.5820/0.5915/ | 0.6703/0.5929/0.5229/0.4587/0.4748/ | 0.7660/0.6930/0.7150/0.6051/0.6184/ |
| 60%          | 0.0641/0.0540/0.0485/0.0463/0.0490/ | 0.0906/0.0752/0.0807/0.0681/0.0682/ | 0.0619/0.0513/0.0464/0.0432/0.0468/ | 0.0913/0.0630/0.0762/0.0641/0.0655/ | 0.0568/0.0485/0.0435/0.0409/0.0427/ | 0.0968/0.0583/0.0713/0.0599/0.0603/ | 0.0517/0.0452/0.0365/0.0327/0.0356/ | 0.0850/0.0513/0.0635/0.0508/0.0566/ | 0.0449/0.0439/0.0291/0.0262/0.0296/ | 0.0691/0.0456/0.0516/0.0415/0.0429/ | 0.0559/0.0486/0.0408/0.0379/0.0407/ | 0.0866/0.0587/0.0687/0.0567/0.0587/ |
Chicago, and the data sparsity is slightly alleviated. In addition, EBPRMF has the optimal F1 value on both two datasets, which shows that EBPRMF proposed in this paper can improve the accuracy of the recommendation list to a certain extent.

### VIII. CONCLUSION

This paper extracts the geographic information and social relationship information of POI respectively, and proposes for the first time that leverage the embedding compression method to integrate the two kinds of information of POI to construct a geographic-social relationship evaluation model. On this basis, the POIs are clustered by spectral clustering algorithm, and multiple clusters with divergence are obtained to ensure the diversity of POI recommendation. EBPRMF is used to learn in each cluster, and the POIs that are closest to the user’s preference are picked up to form a recommendation list and presented to the user. Finally, the effectiveness of EBPRMF is verified by experiments on real datasets.

POI recommendation is currently a valuable research hotspot, and there is a lot of knowledge worth learning in this field. In the next work, we will consider and combine more factors that can affect the recommendation results of POI.

1) **Text information of POI:** The text information of the POI can directly express many characteristics of POI, so we will add the text information of POI to build a Geo-text-social relationship model to make the recommendation list meet the higher dimensional diversity.

2) **Time Factor:** Users will have different POI needs at different times. For example, a person may prefer amusement parks when he/she is young and may prefer movie theaters when he/she grows up. Obviously, the time factor plays an important role in POI recommendation. The next step is to add the above two points to the POI recommendation to enhance the effect of the recommendation.

### REFERENCES

[1] Y. Liu, T.-A.-N. Pham, G. Cong, and Q. Yuan, “An experimental evaluation of point-of-interest recommendation in location-based social networks,” Proc. VLDB Endowment, vol. 10, no. 10, pp. 1010–1021, Jun. 2017.

[2] W. Wang, H. Yin, L. Chen, Y. Sun, S. Sadiq, and X. Zhou, “Geo-SAGE: A geographical sparse additive generative model for spatial item recommendation,” in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), Sydney, NSW, Australia, 2015, pp. 1255–1264.

[3] H. Li, R. Hong, Z. Wu, and Y. Ge, “A spatial-temporal probabilistic matrix factorization model for point-of-interest recommendation,” in Proc. SIAM Int. Conf. Data Mining, Miami, FL, USA, Jan. 2016, pp. 117–125.
[4] S. Zhao, M. R. Lyu, and I. King, “ STELLAR: Spatial-temporal latent ranking model for successive POI recommendation,” in Proc. 30th AAAI Conf. Artif. Intell., 2016, pp. 315–322.

[5] M. Deng, X. Liu, and B. Li, “Diversified recommendation method based on user preference and dynamic interest,” J. Chin. Comput. Syst., vol. 39, no. 9, pp. 2020–2034, 2018.

[6] D. Jiang, K. W.-T. Leung, J. Vosecky, and W. Ng, “Personalized query suggestion with diversity awareness,” in Proc. IEEE 30th Int. Conf. Data Eng., Chicago, IL, USA, Mar. 2014, pp. 400–411.

[7] P. Cheng, S. Wang, J. Ma, J. Sun, and H. Xiong, “Learning to recommend accurate and diverse items,” in Proc. 26th Conf. World Wide Web, Apr. 2017, pp. 183–192.

[8] J. Liu, K. Deng, H. Sun, Y. Ge, X. Zhou, and C. S. Jensen, “Clue-based spatio-textual query,” Proc. VLDB Endowment, vol. 10, no. 5, pp. 529–540, Jan. 2017.

[9] L. Chen, S. Shang, C. S. Jensen, B. Yao, and P. Kalnis, “Parallel semantic trajectory similarity join,” in Proc. IEEE 36th Int. Conf. Data Eng. (ICDE), Apr. 2020, pp. 997–1008.

[10] C. Gao and S. J. Christian, “Querying geo-textual data: Spatial keyword queries and beyond,” in Proc. Int. Conf. ACM SIGMOD, San Francisco, CA, USA, 2016, pp. 2207–2212.

[11] L. Chen, G. Cong, C. S. Jensen, and D. Wu, “Spatial keyword query processing: An experimental evaluation,” Proc. VLDB Endowment, vol. 6, no. 3, pp. 217–228, 2013.

[12] H. Wang, M. Terrovitis, and N. Mamoulis, “Location recommendation in location-based social networks using user check-in data,” in Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL), Orlando, FL, USA, 2013, pp. 364–373.

[13] R. Gao, J. Li, B. Du, X. Li, J. Chang, C. Song, and D. Liu, “Exploiting geo-social correlations to improve pairwise ranking for point-of-interest recommendation,” China Commun., vol. 15, no. 7, pp. 180–201, Jul. 2018.

[14] Z. Pan, L. Cui, X. Wu, Z. Zhang, X. Li, and G. Chen, “Deep potential geo-social relationship mining for point-of-interest recommendation,” IEEE Access, vol. 7, pp. 99496–99507, 2019.

[15] C. Xu et al., “Point-of-interest recommendation algorithm combining location influence,” J. Comput. Appl., vol. 39, no. 11, pp. 3178–3183, 2019.

[16] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, “Exploiting geographical influence for collaborative point-of-interest recommendation,” in Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. (SIGIR), Beijing, China, 2011, pp. 325–334.

[17] K. Lei et al., “Content-based POI recommendation in real-time traffic constraint,” Comput. Eng., vol. 43, no. 10, pp. 147–152, 2017.

[18] L. Huang et al., “Application of content-based recommendation algorithm to traffic signal timing,” Logistics Eng. Manage., vol. 41, no. 10, pp. 124–127, 2019.

[19] X. Zhou, W. Shu, F. Lin, and B. Wang, “Confidence-weighted bias model for online collaborative filtering,” Appl. Soft Comput., vol. 7, no. 5, pp. 1–27, 2017.

[20] C. Yang, X. Yu, Y. Liu, Y. Nie, and Y. Wang, “Collaborative filtering with weighted opinion aspects,” Neurocomputing, vol. 210, no. 19, pp. 185–196, Oct. 2016.

[21] B. Liu, H. Xiong, S. Papadimitriou, Y. Fu, and Z. Yao, “A general geographical probabilistic factor model for point of interest recommendation,” IEEE Trans. Knowl. Data Eng., vol. 27, no. 5, pp. 1167–1179, May 2015.

[22] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, “GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation,” in Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (KDD), 2014, pp. 831–840.

[23] C. Cheng, H. Yang, I. King, and M. R. Lyu, “Fused matrix factorization with geographical and social influence in location-based social networks,” in Proc. 20th Conf. Artif. Intell. Toronto, ON, Canada: AAAI, 2012.

[24] Y. L. Zhao, L. Nie, X. Wang, and T. S. Chua, “Personalized recommendations of locally interesting venues to tourists via cross-region community matching,” ACM Trans. Intell. Syst. Technol., vol. 5, no. 3, p. 50, 2014.

[25] E. H. F. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in Proc. Int. Conf. Knowl. Discovery Data Mining, Portland, OR, USA, 1996, pp. 226–231.

[26] G. Karypis, E.-H. Han, and V. Kumar, “Chameleon: Hierarchical clustering using dynamic modeling,” Computer, vol. 32, no. 8, pp. 68–75, 1999.

[27] S. Gao, K. Janowicz, and H. Couclelis, “Extracting urban functional regions from points of interest and human activities on location-based social networks,” Trans. GIS, vol. 21, no. 5, pp. 446–467, Jun. 2017.

[28] J. He, X. Li, and L. Liao, “Category-aware next point-of-interest recommendation via listwise Bayesian personalized ranking,” in Proc. 26th Int. Joint Conf. Artif. Intell., Melbourne, VIC, Australia, Aug. 2017, pp. 1837–1843.

[29] C. Zou and Z. Chen, “Joint latent factors and attributes to discover interpretable preferences in recommendation,” Inf. Sci., vol. 505, pp. 498–512, Dec. 2019.

[30] L. Huang, Y. Ma, Y. Liu, and A. K. Sangaiah, “Multi-modal Bayesian embedding for point-of-interest recommendation on location-based cyber-physical–social networks,” Future Gener. Comput. Syst., vol. 108, pp. 1119–1128, Jul. 2020.

[31] X. Han et al., “Personalized point-of-interest recommendation in location-based social networks,” Appl. Res. Comput., vol. 36, no. 5, pp. 190–194, 2019.

[32] H. Yin, X. Zhou, B. Cui, H. Wang, K. Zheng, and Q. V. H. Nguyen, “Adapting to user interest drift for POI recommendation,” IEEE Trans. Knowl. Data Eng., vol. 28, no. 10, pp. 2566–2581, Oct. 2016.

[33] Y. Zhang, G. Liu, A. Liu, Y. Zhang, Z. Li, X. Zhang, and Q. Li, “Personalized geographical influence modeling for POI recommendation,” Intell. Syst., vol. 35, no. 5, pp. 18–27, Sep./Oct. 2020.

[34] F. Zhou, R. Yin, K. Zhang, G. Trajcevski, Z. Zhong, and J. Wu, “Adversarial point-of-interest recommendation,” in Proc. World Wide Web Conf. (WWW), 2019, pp. 3462–3468.

[35] X. Meng et al., “A diversified and personalized recommendation approach based on geo-social relationships,” Chin. J. Comput., vol. 42, no. 11, pp. 2574–2590, 2019.

[36] J. Zhao, Y. Gao, C. Ma, P. Jin, and S. Wen, “On efficiently diversified top-k geo-social keyword query processing in road networks,” Inf. Sci., vol. 512, pp. 813–829, Feb. 2020.

[37] H. Borges, R. Guibert, O. Permiakova, and T. Burger, “Distinguishing between spectral clustering and cluster analysis of mass spectra,” J. Proteome Res., vol. 18, pp. 1–4, Apr. 2019.

[38] Z. Wu and R. Leahy, “An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, no. 11, pp. 1101–1113, 1993.

[39] X. Zhang, Matrix Analysis and Applications. Beijing, China: Tsinghua Univ. Press, 2015, pp. 528–540.

[40] G. Hu and X. Dai, “Integrating reviews into personalized ranking for cold start recommendation,” in Proc. Conf. Knowl. Discovery Data Mining Pacific-Asia, 2017, pp. 708–720.

[41] F. Zhao, Y. Shen, X. Gui, and H. Jin, “SDBPR: Social distance-aware Bayesian personalized ranking for recommendation,” Future Gener. Comput. Syst., vol. 95, pp. 372–381, Jun. 2019.

[42] Y. Fang, Y. Guo, and J. Lan, “A Bayesian probabilistic matrix factorization algorithm based on logistic function,” J. Electron. Inf. Technol., vol. 42, no. 3, pp. 715–720, 2014.

XIAOFU MENG was born in 1981. He received the Ph.D. degree from Northeastern University, China, in 2010. He is currently a Full Professor and a Ph.D. Supervisor with Liaoning Technical University, China. His research interests include spatial data management, recommender systems, and web database query.

JINFENG FANG was born in 1992. She received the M.S. degree from Liaoning Technical University, China, where she is currently pursuing the Ph.D. degree. Her research interests include spatial data analysis, POI recommendation, and web database query.