



SQPMF: successive point of interest recommendation system based on probability matrix factorization

Jie Wang¹ · Zhiyi Huang² · Zhaobin Liu¹

Accepted: 24 November 2023 / Published online: 18 December 2023
© Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

With the development of Location-Based Social Networks, successive Point Of Interest (POI) recommendation systems have become a hot spot in the field of recommendation systems. Successive POI recommendation systems suggest to users new and interesting places to visit. However, in real-life POI recommendation, there are often a small number of users facing a huge number of POIs. Traditional successive POI recommendation methods are not capable to deal with the large sparse datasets, as they only consider the simple relationship between users and POIs. They do not use the context information of users and POIs, which can enable better recommendation results. To utilize the context information, this paper proposes a novel successive POI recommendation method, SQPMF, which integrates user personal preferences, user social relationships and POI transition relationships into the system for accurate recommendation of the next POI. Our experimental evaluation using three real-life datasets, Gowalla, Foursquare and Brightkite, shows that our method SQPMF consistently outperforms all state-of-the-art methods in recommendation of successive POIs. Compared with other methods, SQPMF improves Precision, Recall and F1-score by an average of at least 6.1%, 5.8% and 5.7% respectively on three publicly available datasets.

Keywords Successive POIs · Recommendation system · Probability matrix factorization · Location-based social networks

1 Introduction

In recent years, with the rapid development of the Internet, various Location-Based Social Networks (LBSN) such as Twitter, Facebook, Weibo, and Meituan have emerged. The rapid development of these social networks has brought many conveniences to people's lives and greatly improved people's quality of life. People can use their mobile devices to share check-in information in various situations in their lives. Using the users' check-in information to recommend Points Of Interest (POI) is also a popular development trend in recent years [1–3]. LBSN can obtain valuable information by collecting users' check-in data. A POI refers to a place

where a user checks in, such as a shopping mall, a cafe, etc. POI recommendation can effectively alleviate the problem of overloading location information and improve users' personalized experience. At the same time, it helps businesses to tap potential customers and improve business benefits. Therefore, POI recommendation has become an important service in LBSN and one of the core research directions in the field of LBSN and recommendation systems [4, 5]. An important challenge that needs to be solved in POI recommendation systems is the problem of data sparsity. There are thousands of POIs in a LBSN, but users can only go to a limited number of POIs, which will result in very sparse check-in data [6, 7]. In real-life, people check in at different locations anytime and anywhere, leaving a large amount of feedback data. But for hundreds of millions of POIs, there is very little explicit data available to users [8–10]. In order to alleviate the above problem, most recommendation systems use Collaborative Filtering (CF) to recommend items or POIs. Most of the current research on POI recommendation systems only consider the check-in relationship between users and POIs, ignoring the correlation between POIs that users visit in sequence. In fact, there is a strong correlation between a user's current POI and their next POI to visit. For

✉ Zhaobin Liu
zhbliu@dlnu.edu.cn

Jie Wang
wangjie0606@dlnu.edu.cn

Zhiyi Huang
hzy@cs.otago.ac.nz

¹ College of Information Science and Technology,
Dalian Maritime University, Dalian 116026, China

² University of Otago, Dunedin 9054, New Zealand

example, after going shopping in a mall, users usually go to a restaurant to eat instead of traveling abroad. Therefore, in the process of building a POI recommendation model, it is necessary to recommend the next POI to a user based on the user's current POI [11–13].

Additionally, in real life, personal preferences have a great impact on users' behaviors. The number of visits to each POI shows a user's personal interest preferences for the POIs [14–16]. On the other hand, a user's personal preferences are usually influenced by their friends because of similar preferences. Therefore, if a user has not visited a place yet, the personal preferences of their social friends could provide useful information for the recommendations [17–19].

Most existing POI recommendation systems only consider the partial relationship between users and POIs, which is not comprehensive enough. Therefore, in this paper, a successive POI recommendation method, called SQPMF, is proposed. It integrates the user's personal preferences, the user's social relationships and the transition relationships between successive POIs to improve the accuracy of the recommendation system. In order to mine the features of user's social relationships, we study the explicit trust relationships and implicit trust relationships between users in SQPMF. We use the habit similarity between users and the sequence similarity between users to further mine the hidden information behind the user's rating on POIs [20–22]. In SQPMF, we fuse together the transition frequency between POIs, the geographic distance between POIs, and the popularity of destination, which also have a large impact on the recommendation results of the next POI. We use Probability Matrix Factorization (PMF) to model and solve the user's social relationships and the POI transition relationships in SQPMF. PMF is a popular method in building recommendation systems. It performs well on large sparse datasets. In recent years, PMF has attracted extensive attention of researchers with its good accuracy and scalability [23, 24] for solving the recommendation problem. Experimental results show that SQPMF can significantly improve the accuracy of successive POI recommendations. The main contributions of this paper are as follows.

- (1) We propose a new successive POI recommendation method, SQPMF, that utilizes the user's personal preferences, the user's social relationships and the transition relationships between successive POIs, to improve the accuracy of successive POI recommendations.
- (2) We prove that SQPMF can most likely produce accurate recommendations of POIs based on Bayes' Theorem and the assumptions of PMF.
- (3) Several experiments are conducted with real datasets. The experimental results show that SQPMF has much better performance than the existing recommendation methods. It improves Precision, Recall and F1-score by

an average of 6.1%, 5.8%, 5.7% respectively on three publicly available datasets.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 introduces the notation and definitions used in this paper, the details of the proposed SQPMF method, and the derivation of related formulas. Section 4 presents the experimental setup, results, analysis and evaluation. Finally, Section 5 concludes this paper.

2 Related work

In the field of POI recommendation, in order to alleviate the problem of data sparsity, most methods adopt the Collaborative Filter (CF) methods. The CF methods are mainly divided into memory-based methods and model-based methods [25, 26]. Memory-based methods have achieved great success in the POI recommendation system, but many memory-based methods suffer from data sparsity and cold start problems. However, with the rapid development of machine learning technology, model-based methods have emerged. The well-known models are Bayesian models, clustering models, and etc. Model-based methods can avoid the shortcomings of memory-based methods. Among them, the Matrix Factorization (MF) is a typical model-based method. At present, the basic MF assumes a linear relationship between users and POIs and has good scalability and flexibility. Since it can alleviate the problem of data sparsity to a certain degree, MF is widely used. Rahmani et al. [24] proposed a POI recommendation method based on a local geographical model, and integrated this model into MF. This improved the performance of recommendation to some extent. Seyedhoseinzadeh et al. [25] proposed a new framework based on the ideas of CF and MF. It modeled the recommendation problem of POI based on users and POIs, and partially addressed the problem of data sparsity.

However, the basic MF method does not take users, POIs, and contextual features into account for modeling. This shortcoming causes the loss of useful information and limits the accuracy of the recommendation results. In successive POIs recommendation systems, a small number of users are often mapped to a large number of POIs. With hundreds of millions of POIs, MF is faced with huge computational overhead. Thus, MF is hard to deal with large datasets. In order to alleviate this problem of MF, Mnih et al. [26] proposed the PMF model. PMF is similar to MF, which is based on the low-rank approximation of the matrix, except PMF adds Bayes' Theorem to MF. It adopts the probability linear setting with Gaussian observation noise, and decomposes the user rating matrix into two low-rank user feature matrices and two item feature matrices with the most likelihood. The authors of [27, 28] have shown that PMF performs well in large-scale,

sparse datasets. Among them, Davtalab et al. [27] proposed a method called SSTPMF, which is based on PMF. SSTPMF utilized information about users and POIs to improve the accuracy of POI recommendations to a certain extent. However, the authors failed to use the relationships of users and POIs sufficiently. For example, when calculating POI similarity, SSTPMF only mapped the POI category score to 0 or 1, which greatly affected the accuracy of the results. In addition, in mining the social relationships of users, the authors only used the Jaccard measure representing user social similarity. Jaccard measure is generally used to reflect whether the user will choose a POI, but cannot reflect the specific preferences of the user related to the POI. In fact, the scores of social relationships can be expressed more comprehensively through using other available information. For example, by considering information such as users' habit similarity and sequence similarity, we can further understand the users' potential preferences for POI, and thus better improve the accuracy of recommendations. Therefore, our SQPMF method is proposed to address these issues, and to improve the performance of the recommendations for successive POIs.

Though explicit feedback such as ratings can clearly express the likes and dislikes of users, POI recommendation systems only have users' check-in information of the visited POIs, which do not directly express the user's preferences. This kind of check-in information is called implicit data [29–31]. In POI recommendation, only using the explicit data would result inaccurate recommendations. It is necessary to integrate implicit data of users and POIs in MF. In order to address this problem, many works [1, 8, 21, 32–34] considered incorporating rich implicit information into POI recommendation systems.

For example, in LSBN, there are a lot of connections between users, which play an important role for accurate recommendations. Guo et al. [1] proposed a trust-based MF model, which comprehensively considered the explicit and implicit trust relationships between users and achieved better recommendation results. Ma et al. [34] integrated the users' trust relationships in PMF and used the users' social network information and rating records to solve the problems of data sparsity and poor prediction accuracy. They showed the effectiveness of the trust relationships in improving the recommendation results. Qian et al. [21] proposed a personalized recommendation system based on PMF that combined users' personal interests, similarity of interpersonal interests, and interpersonal influence, which further improved the accuracy of recommendation. Zhou et al. [8] considered the strength of user relationships using data-driven methods to improve POI recommendations. It defined user relationships based on the analysis of user check-in behaviors. They embedded the user's social connections into a spatiotemporal model of POI recommendations.

It is worth noting that the recent application of Knowledge Graphs (KGs) has been proved to be effective in addressing the problem of data sparsity. It is mainly used to learn the characteristics of users and POIs from the check-in records of users and their friends. Chen et al. [35] used KGs to design a new spatiotemporal transition relationship and jointly trained and modeled the next POI recommendation system in an end-to-end manner. Hu et al. [36] proposed a translation-based KG-enhanced multi-task learning framework (TransMKR) for POI recommendation. TransMKR improved POI recommendation based on KGs, quantified the relationships between POIs and their attributes, and thus alleviated the problem of data sparsity. Though these methods greatly improved the recommendation quality, there is still much room for the application of KGs. For example, the current KGs cannot reflect the dynamic movement of the POIs with static entities and relationships.

Geographical influence is also an important factor to consider in POI recommendation, since a user's preference for a POI is largely influenced by the geographical location of the POI. Lian et al. [37] proposed a GeoMF model, which incorporated the spatial clustering phenomenon of users movements into the factorization model and enhanced the potential factors of users and POIs in the factorization model by using the activity area vector of users and the influence area vector of POIs. This research attracted the attention of many researchers. Rahmani et al. [16] well considered the geographical factors of POIs and combined time, social and other information to develop a POI context fusion method based on linear regression. This method could effectively find the optimal context combination from the historical interactions of each user or user group, and improved the accuracy of POI recommendation.

Unlike the traditional POI recommendations, the next POI recommendation focuses more on sequence dependencies. Many studies used Markov chain to recommend the next possible POI. Models based on Markov mainly used transition matrixes to predict the probability of the next behavior of a user. For example, PRME (Personalized Ranking Measurement Embedding) [29] was proposed to use user preferences, POI sequential information and geographic influence. Although multistage Markov chain can improve the accuracy of recommendation results to some extent, the computational complexity increases exponentially with the increase of Markov order. The high computational complexity makes the Markov models difficult for practical applications.

Recently, some researchers found that user preferences change over time, especially the current check-in order, which has a significant impact on POIs. Therefore, in order to improve recommendation performance, it is necessary to consider both long-term and short-term preferences of users [38–40]. For example, Liu et al. [38] proposed a model based

on Graph Neural Network (GNN), which integrated users' long-term and short-term preferences to comprehensively represent dynamic preferences. Long Short Term Memory (LSTM), as a variant of Recurrent Neural Network (RNN), is also widely used to recommend the next POI. Liu et al. [39] proposed a Real-Time Preference Mining (RTPM) model based on LSTM, which could dynamically mine users' long-term and short-term preferences, filtered out unpopular POIs, and achieved better recommendation results. However, the GNN/LSTM based methods cannot model the relationships between two discontinuous POIs.

Table 1 summarizes the related methods and their advantages/disadvantages in POI recommendation.

In summary, existing studies have not considered the individual behavioral habits of users. In addition, the previous works only considered the user's partial social relationships in their models. Although these models could alleviate the problem of data sparsity to some extent, they performed poorly for extremely sparse user-POI matrices. Likewise, for POI transition relationships, previous models only considered the geographic distance or the sequential influence of successive POIs, which are not sufficient for accurate recommendations.

In contrast, our SQPMF model not only considers users' sequential preferences, but also integrates their real-time habits and preferences, which helps to understand their current intentions. Therefore, our proposed SQPMF model can provide better recommendations for next POIs by exploiting more implicit information of users' relationships and POI transition relationships.

3 Method

We propose the SQPMF model by exploiting the users' personal preferences, the users' social relationships and the transition relationships between successive POIs. Through model inference, we propose an objective function to learn the users' social relationships and POI transition relationships. We then integrate the learned implicit information into users' ratings of POIs. We use the Stochastic Gradient Descent (SGD) method to minimize our objective function in order to improve the accuracy of the users' rating of POIs, according to which a ranked list of recommendations for the users' next POIs can be obtained. For the convenience of

Table 1 Related research

Methods	Descriptions	Advantages	Disadvantages
PMF [26]	It proposes a PMF model, incorporating a Bayesian perspective on the basis of MF.	It is suitable for large-scale sparse matrices.	It does not take into account the potential relationship between users and POIs.
SSTPMF [27]	It is based on a PMF model to mine information between users and POIs.	It takes into account user similarity and POI similarity.	It cannot accurately calculate user similarity and POI similarity.
TrustSVD [1]	It merges user trust relationships into SVD++.	It considers and handles trust relationships between users.	It sets trust intensity equally for all users, which leads to inaccurate recommendations.
Sorec [34]	It is based on PMF model and incorporates a user trust relationship matrix.	It takes into account trust relationships between users and is suitable for large-scale datasets.	It does not consider other social relationships between users and does not consider the POI transition relationships.
GeoMF [37]	It makes recommendations based on the geographic model of the user's activity area and POI influence area.	It takes into account the geographical influence between users and POIs.	It does not take into account other implicit information between users and POIs, and the impact of user social relationships.
PRME [29]	It utilizes user preferences and sequential transition of POIs.	It takes into account context information such as user preferences and transition relationships between POIs.	It is based on a multi-stage Markov chain and is difficult to scale to large-scale datasets.
RTPM [39]	It proposes a real-time preference mining model based on LSTM.	It takes into account users' long-term and short-term preferences.	It does not fully explore the potential information between users and POIs.

Table 2 Symbols and definitions used in this paper

Symbols	Definitions
W	Feature dimension
M	The number of users
N	The number of POIs
U	User feature matrix
V	POI feature matrix
F	Social feature matrix of users
P	Relevance feature matrix of POIs
R_{ij}	User i 's rating for POI j
R'_{ij}	User i 's predicted rating for POI j
R^z_{ij}	Predicted score of user i for the next POI z when at POI j
S_{if}	Social score of user i and user f
S'_{if}	Predicted social score of user i and user f
Q_{jk}	Transition score of POI j to POI k
Q'_{jk}	Predicted transition score of POI j to POI k
Ψ	Objective function

expression and explanation, the symbols used in our model are listed in Table 2.

3.1 Overall framework

As mentioned above, MF has been widely used in POI recommendation. Due to the shortcomings of MF, such as unable to integrate the context information, unable to process large-scale datasets, PMF has been proposed.

In order to better illustrate our proposed SQPMF, this section makes an overall comparison between SQPMF and MF.

As shown in Fig. 1, MF decomposes the initial sparse matrix $R(M \times N)$ into a user feature matrix $U(M \times W)$ and a POI feature matrix $V(W \times N)$, where M is the number of users, N is the number of POIs, and W is the vector dimension, i.e., the number of features [24, 25].

In Fig. 1, in the initial rating matrix R , users have rated some POIs. For example, user A rated POI Y as 3.80. The question mark means that the user has never visited the POI, e.g. user B has never visited POI Y . In practice, we cannot set the value here as zero, because zero is also a rating, meaning the user does not like the POI. Since the user has never visited the POI, we don't know whether the user likes the POI or not. Through MF, the matrix R , is decomposed into the user feature matrix U and the POI feature matrix V . By multiplying U and V , the predicted rating matrix R' can be calculated. In the predicted rating matrix R' , the missing values (?) of R are filled by the MF method. Therefore, with MF, the recommendation problem is transformed into the matrix factorization problem that aims to reduce the discrepancy between R and R' .

However, for POI recommendation, MF cannot effectively mine the information of users and POIs, which has poor recommendation performance for extremely sparse user-POI matrix.

Considering these shortcomings, this paper integrates the users' social relationships and the POI transition relationships in our SQPMF model to improve the recommendation results. Our experimental results show that all the information we integrate in SQPMF have a positive impact on the successive POI recommendations. Figure 2 shows the model diagram of SQPMF.

Comparing between Figs. 1 and 2, we can see that, based on the MF model, SQPMF combines the users' social relationships and the POI transition relationships. The matrices S and S' represent the initial social relationship matrix and the predicted social relationship matrix, respectively. Integrating more information of users' social relationships into the matrix S will help us to mine more potential preferences of users and help us make better recommendations. Therefore, we integrate the trust relationship, the habit similarity and the sequence similarity into the users' social relationship matrix. It can be seen in matrix S that integrating these factors into the recommendation system can better update the user feature matrix U . It is worth noting that in the initial dataset, a user's social relationship with another user is only

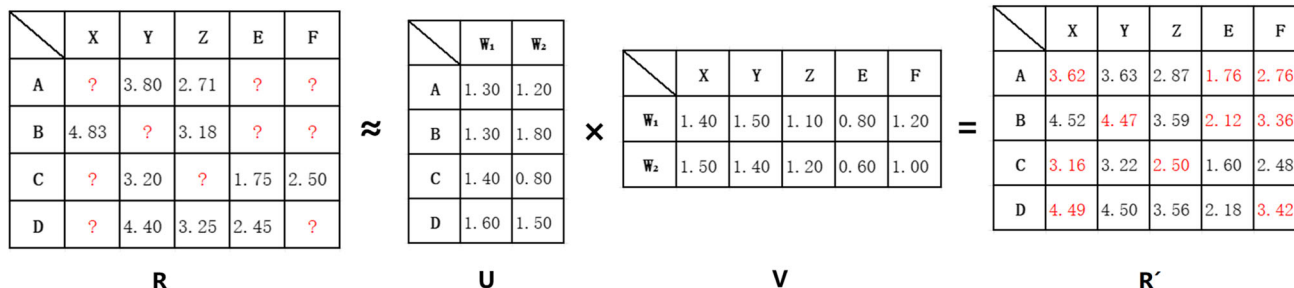


Fig. 1 Matrix factorization model

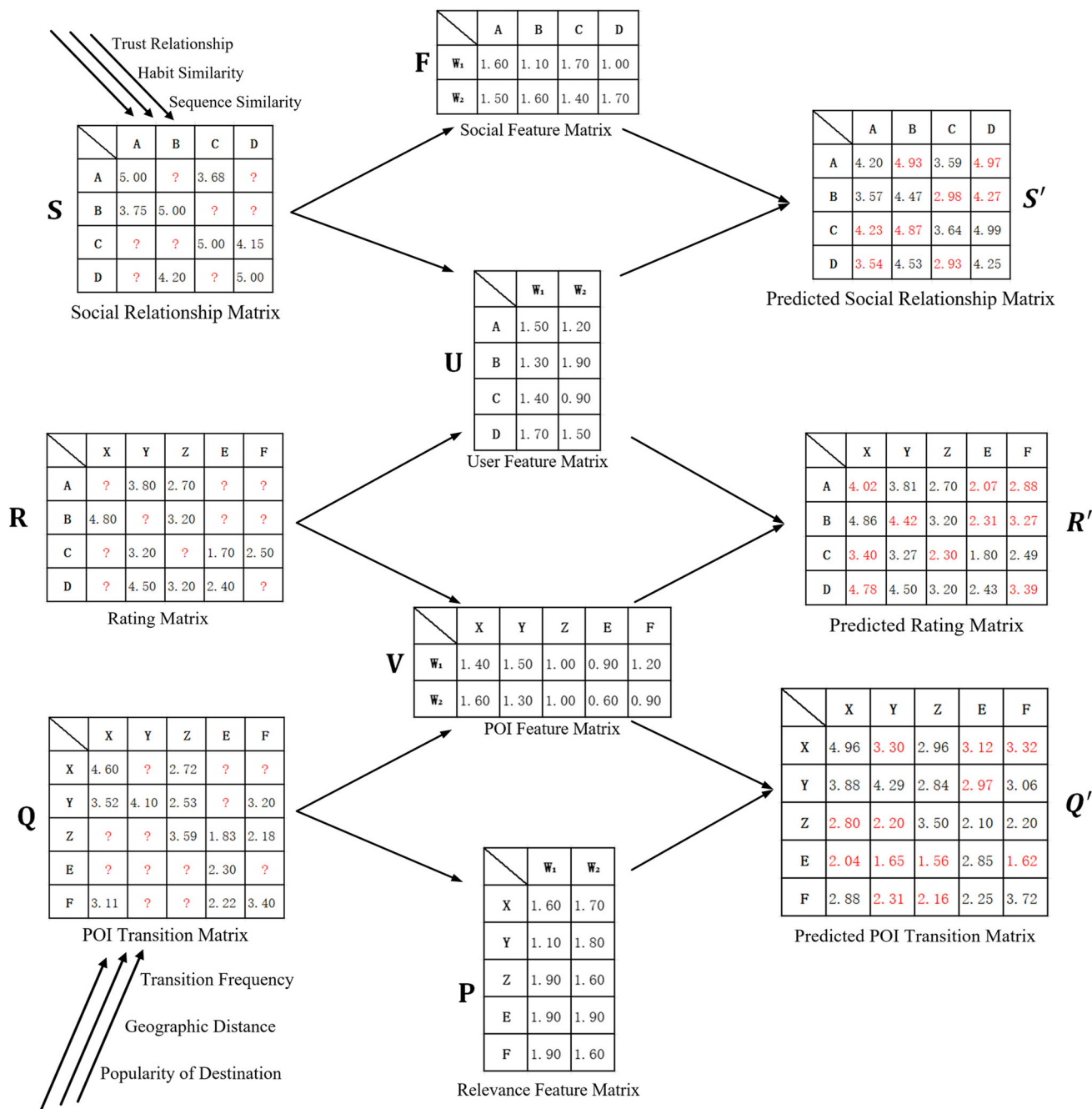


Fig. 2 Model diagram of SQPMF

represented by 0 or 1, which cannot well distinguish the strength of social relationships between users [34]. Therefore, we propose several different calculation formulas to model a user’s social relationship, which better fits the real-life situation. The strength of social relationships between users affects the final recommendation results. After decomposing the social relationship matrix S , it can be divided into a user feature matrix U and a social feature matrix F .

On the other hand, the matrices Q and Q' represent the initial POI transition matrix and the predicted POI transition matrix, respectively. Similarly, considering real-life situations, we want to fuse as much related information as possible in the recommendation of POIs. Most previous works only considered partial relationships between POIs, which was not comprehensive enough. In this paper, the transition frequency, the geographic distance, and the popularity

of destination are considered in the POI transition matrix. This will enable us to learn sufficient information about POIs and help us better make up for the shortcomings of the sparse matrix. The POI transition matrix Q can be divided into POI feature matrix V and relevance feature matrix P of POIs after decomposing.

By continuously decomposing and learning the initial sparse matrix, using the information learned from the social relationship matrix S and the POI transition matrix Q , the user feature matrix U and the POI feature matrix V are continuously updated. At the same time, the social feature matrix F and the relevance feature matrix P are constantly updated, so that the discrepancy between the matrix S and the matrix S' and the discrepancy between the matrix Q and the matrix Q' are continuously reduced. The update process iterates until the discrepancy between the initial score matrix R and the predicted score matrix R' is smaller enough. As a result, the values of R' in Fig. 2 are closer to the values of R than those of R' in Fig. 1. This example illustrates that our method of integrating the social relationship matrix S and POI transition matrix Q into SQPMF is very effective. The following section will introduce our mathematical formulas and detailed explanations. Our experimental results will show that SQPMF is more suitable for large-scale successive POI recommendations where the user-POI matrix R is extremely sparse.

3.2 Metric for user social relationship

In the field of POI recommendation, explicit and implicit trust relationships are two different ways of expressing relationships between users. The essential difference between explicit trust relationships and implicit trust relationships lies in their different expressions. Explicit trust relationship refers to the level of trust that users have towards other users or POIs expressed through clear ratings or feedback. For example, on social media platforms, users can like, comment on, or share their content with friends, which directly express trust and recognition towards their friends. Implicit trust relationship refers to the indirect expression of a user's level of trust in other users through user behavior or implicit feedback. These behaviors may include user browsing history, purchase history, click behavior, etc. By analyzing users' implicit feedback information, it is possible to infer their level of interest in a particular POI and their trust in other users. Usually the explicit trust relationships we can obtain in real life are limited. Therefore, considering both explicit and implicit trust relationships can effectively represent the trust relationships and thus helps improve the accuracy of recommendations.

According to the previous research [41, 42], a user's social relationship has a significant impact on the user's activity trajectories. Recommendation methods based on users' social relationships can effectively address the data sparsity

problem and improve recommendation results [43, 44]. Traditionally the social relationship between user i and user f is represented by 0 or 1. If user i trusts user f , then the social relationship from user i to user f is represented by 1; otherwise, it is 0. However, expressing the user's social relationship with 0 or 1 cannot well represent the strength of social relationships between users.

Considering the real-life social relationships, our SQPMF model integrates both the explicit trust relationships and implicit trust relationships in successive POI recommendation, to improve the accuracy of the recommendations.

In real life, expressing a user's social relationship with 0 or 1 hides many details, such as the trust relationship, the habit similarity and the sequence similarity, between users. We use a novel social score considering the above factors to represent the degree of social relationship. The social score is expressed in (1):

$$S_{if} = \eta S_{if}^{imp} + \mu S_{if}^{hab} + (1 - \eta - \mu) S_{if}^{seq} \quad (1)$$

where S_{if}^{imp} represents the trust degree between users, S_{if}^{hab} represents the habit similarity, S_{if}^{seq} represents the sequence similarity, $\eta, \mu \in [0, 1]$ represent the weight of different attributes, respectively.

Note that, in real life, a user's social relationship is usually related rather than symmetrical. Therefore, the matrix S is asymmetric, that is, the social score from user i to user f is not the same as the social score from user f to user i .

Below are the details of S_{if}^{imp} , S_{if}^{hab} and S_{if}^{seq} used in (1).

In real life, a user has many friends, some of them have similar habits to the user, but others may have different habits from the user. These details could affect the accuracy of the recommendations. In order to learn as much relevant information of users as possible in the recommendation, and to mine the implicit information in the users' relationships, we propose to integrate the habit similarity in a successive POI recommendation, as shown in (2).

$$S_{if}^{hab} = \sum_{i=1}^M \sum_{f \in F} S_{if}^* \times \|U_{ij} - U_{fj}\|_F^2 \quad (2)$$

where U_{ij} and U_{fj} represent the difference from the ratings of user i to user f for POI j , S_{if}^* represents the original social relationship of user i to user f , and the larger the value of S_{if}^{hab} , the higher the habit similarity from user i to user f .

In successive POI recommendation, each user has its own historical access sequence. The users' access to POIs can describe the users' preference transition information for POIs. If the overlap of the historical access sequence is high between two users, we think that the sequence similarity is

also high. Therefore, we calculate the sequence similarity in (3):

$$S_{if}^{seq} = \frac{seq_{if}^2}{seq_i seq_f} \quad (3)$$

where seq_i and seq_f represent the total number of sequences of user i and user f , respectively. seq_{if} represents the number of common sequences between user i and user f . The larger the value of S_{if}^{seq} , the higher the sequence similarity from user i to user f .

Considering the influence of the implicit trust relationship on the trust relationship, it is necessary to improve the representation of the trust degree. In location-based social networks, we believe that if user i trusts many users, the confidence of user i 's trust value will decrease. However, if user f is trusted by many users, the confidence of user f 's trust value should increase. We use (4) to express trust degree [31]:

$$S_{if}^{imp} = \sqrt{\frac{d^-(U_f)}{d^+(U_i) + d^-(U_f)}} \times S_{if}^* \quad (4)$$

where $d^+(U_i)$ represents the outdegree of user i , which is the number of users trusted by user i . $d^-(U_f)$ represents the indegree of user f , which is the number of users who trust user f . S_{if}^* represents the original social relationship between user i and user f (0 or 1). The larger the value of S_{if}^{imp} , the higher the degree of trust from user i to user f .

It should be noted that the above formula is based on directed networks. Among the datasets used in our experiment, Gowalla uses an undirected friendship network, Foursquare uses a directed friendship network, and Brightkite uses a hybrid network where undirected edges are used when there is a friendship in both ways. All the original datasets downloaded from their websites have been already represented as directed graphs, which is easily applicable to (4). For example, in Gowalla's original dataset, three columns of data were used to represent the original trust relationship between users, namely the user ID, the user's social friend ID, and their relationship (1 represents trust, 0 represents distrust). If there is an undirected edge connecting user 2 and user 4, two records are present in the dataset to represent their trust relationship as below:

2 4 1
4 2 1

However, in cases where the trust relationships between users are represented by an undirected graph, the network can be converted into a directed graph. In the conversion, each undirected edge is transformed into two directed edges. For each user or node in the converted directed graph, the indegree of the user equals the outdegree of the user in such cases. After the conversion, (4) is applicable to undirected graphs to calculate implicit trust relationships.

3.3 Metric for POI sequence transition

Likewise, many previous works only utilize partial information between POIs when expressing the sequence correlation between POIs. However, such practice loses useful information between POIs, such as the transition frequency, the geographic distance, and the popularity of destination. Considering these factors, we propose (5) to evaluate the transition score:

$$Q_{jk} = \beta Q_{jk}^{fre} + \gamma Q_{jk}^{geo} + (1 - \beta - \gamma) Q_{jk}^{pop} \quad (5)$$

where Q_{jk}^{fre} represents the transition frequency score between POIs, Q_{jk}^{geo} represents the geographic distance score between POIs, Q_{jk}^{pop} represents the popularity score of destination, $\beta, \gamma \in [0, 1]$ represent the weight of different attributes, respectively.

Likewise, due to the asymmetry of the sequence of POIs visited by users, the transition score from POI j to POI k is not completely equivalent to the transition score from POI k to POI j . So the matrix Q is also asymmetric.

Below are the details of Q_{jk}^{fre} , Q_{jk}^{geo} , and Q_{jk}^{pop} in (5).

Considering the continuity of the sequence between the POIs visited by the users, we propose to use the continuity between the POIs to improve the accuracy of the recommendation. Since a user's check-in behavior is affected by the user's personal living habits and the transition relationships between the POIs, there is a strong correlation between some sequences of POIs [41]. By processing and studying the sequence changes between POIs in the datasets, we can calculate the transition frequency, and obtain the transition frequency score in (6):

$$Q_{jk}^{fre} = \sum_{i=1}^M N_i(j, k) \quad (6)$$

where $N_i(j, k)$ represents the number of POI transitions for user i from POI j to POI k . Q_{jk}^{fre} represents the total number of POI transitions from POI j to POI k for all users. The larger the value of Q_{jk}^{fre} , the higher the transition frequency score at the given time. Due to the large range of the values, for the convenience of calculation, we use the normalization function $f(x) = (x - Q_{\min}) / (Q_{\max} - Q_{\min})$ to map the value of Q_{jk}^{fre} to the range of $[0, 1]$.

According to Tobler's first law of geography, "Everything is related to other things, but things that are near are more closely related than things that are far away", the probability of users visiting POIs is inversely proportional to the geographic distance in real life. POIs frequently visited by a user are likely to be geographically close, or the POIs frequently visited by the same user are relatively close to each other. Therefore, we calculate the geographic distance score

to describe the users' geographic preference. We use the Haversine formula [45] to calculate the distance between POI j and POI k , as shown in (7).

$$d_{jk} = \sin^{-1} \sqrt{\left(\frac{\sin(\Delta\varphi_{jk})}{2}\right)^2 + \cos\varphi_j \times \cos\varphi_k \times \left(\frac{\sin(\Delta\omega_{jk})}{2}\right)^2} \tag{7}$$

where φ_j and φ_k refer to the latitude of POI j and POI k in radians, and ω_j and ω_k refer to the longitudes of POI j and POI k in radians. Therefore, the geographic similarity of POI j and POI k is shown in (8):

$$Q_{jk}^{Geo} = \frac{1}{1 + (d_{jk} \times r)} \tag{8}$$

where d_{jk} has been calculated above, and r refers to the radius of the earth. The higher the value of Q_{jk}^{Geo} , the higher the geographic distance score between POI j and POI k .

Popularity refers to the degree to which a POI is liked by users. In real life, under the same conditions, users tend to choose a POI with higher popularity to visit. Therefore, using the popularity of destination can also effectively improve the accuracy and quality of recommendation. Therefore, considering the popularity factor of the next destination in the successive recommendation system, we use (9) to calculate:

$$Q_{jk}^{Pop} = \frac{N(j)}{\sum_{j' \in V} N(j')} \tag{9}$$

where $N(j)$ refers to the number of times the POI j has been visited, V refers to the set of all POIs, and j' is any POI in V . The higher the value of Q_{jk}^{Pop} , the higher the popularity of the destination.

3.4 Proposed SQPMF model

As mentioned above, we transform the recommendation problem into an optimization problem that reduces the discrepancy between the user's actual rating and the predicted rating of the POIs. As shown in Fig. 2, our SQPMF model uses two additional matrices S and Q to represent the social relationship matrix and the POI transition matrix [43, 44]. S and Q are created as shown previously and can help us find more accurate prediction of users' ratings on POIs, i.e., minimizing the discrepancy between the matrices R and R' in Fig. 2. In SQPMF, we are trying to minimize our objective function shown in (10), which includes the discrepancy between R and R' , S and S' , Q and Q' . It should be noted that we use the sigmoid function $g(x) = 1/(1 + e^{(-x)})$ to process the value of each element in the prediction matrix R' , S' and

Q' . The reason for this is easy to explain: it maps real-valued input into a continuous linear function that is monotonically increasing on $[0,1]$ and can handle any number of input values. In this way, U is informed by S and V is informed by Q . This means the users' social relationships and POI transition relationships are integrated into our SQPMF model to generate the predicted user rating matrix R' for POI recommendation. The objective function used in SQPMF is shown as (10). The derivation process of the proposed objective function will be shown in the next section.

$$\begin{aligned} \Psi(R, S, Q, U, V, F, P) = & \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^N I_{ij}^R ((R_{ij} - g(U_i^T V_j))^2) \\ & + \frac{\lambda_S}{2} \sum_{i=1}^M \sum_{f=1}^M I_{if}^S ((S_{if} - g(U_i^T F_f))^2) \\ & + \frac{\lambda_Q}{2} \sum_{j=1}^N \sum_{k=1}^N I_{jk}^Q ((Q_{jk} - g(V_j^T P_k))^2) \\ & + \frac{\lambda_U}{2} \sum_i \|U\|_F^2 + \frac{\lambda_V}{2} \sum_j \|V\|_F^2 \\ & + \frac{\lambda_F}{2} \sum_f \|F\|_F^2 + \frac{\lambda_P}{2} \sum_k \|P\|_F^2 \end{aligned} \tag{10}$$

Where the first three items are the squared errors of R and R' , squared errors of S and S' , squared errors of Q and Q' , and $\lambda_U = \frac{\sigma_R^2}{\sigma_U^2}$, $\lambda_V = \frac{\sigma_R^2}{\sigma_V^2}$, $\lambda_F = \frac{\sigma_R^2}{\sigma_F^2}$, $\lambda_P = \frac{\sigma_R^2}{\sigma_P^2}$, $\lambda_S = \frac{\sigma_R^2}{\sigma_S^2}$, $\lambda_Q = \frac{\sigma_R^2}{\sigma_Q^2}$ are the regularization parameters. $\|\cdot\|_F^2$ represents the Froberius norm of the matrix, which can be calculated by the sum of the squares of the absolute values of the elements of the matrix. Its main function is to reflect the discrepancy between the real matrix and the predicted matrix.

This model combines the original PMF model with the social relationship matrix S and the POI transition matrix Q to learn, making full use of the relevant information of users and POIs. By iteratively minimizing the objective function, a better recommendation of POIs to users can be achieved.

We use the Stochastic Gradient Descent (SGD) to minimize our objective function [26–28], as shown in (11), (12), (13) and (14).

$$\begin{aligned} \frac{\partial \Psi}{\partial U_i} = & \sum_{j=1}^N I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - R_{ij})V_j \\ & + \lambda_S \sum_{f=1}^M I_{if}^S g'(U_i^T F_f)(g(U_i^T F_f) - S_{if})F_f \\ & + \lambda_U U_i \end{aligned} \tag{11}$$

$$\begin{aligned} \frac{\partial \Psi}{\partial V_j} &= \sum_{i=1}^M I_{ij}^R g'(U_i^T V_j)(g(U_i^T V_j) - R_{ij})U_i \\ &+ \lambda_Q \sum_{k=1}^N I_{jk}^Q g'(V_j^T P_k)(g(V_j^T P_k) - Q_{jk})P_k \\ &+ \lambda_V V_j \end{aligned} \tag{12}$$

$$\frac{\partial \Psi}{\partial F_f} = \lambda_S \sum_{i=1}^M I_{if}^S g'(U_i^T F_f)(g(U_i^T F_f) - S_{if})U_i + \lambda_F F_f \tag{13}$$

$$\frac{\partial \Psi}{\partial P_k} = \lambda_Q \sum_{j=1}^N I_{jk}^Q g'(V_j^T P_k)(g(V_j^T P_k) - Q_{jk})V_j + \lambda_P P_k \tag{14}$$

Iterating on (11), (12), (13) and (14) can keep updating the user feature matrix U , the POI feature matrix V , the social feature matrix F and the relevance feature matrix P , until the objective function converges.

$$U_i = U_i - \alpha \frac{\partial \Psi}{\partial U_i} \tag{15}$$

$$V_j = V_j - \alpha \frac{\partial \Psi}{\partial V_j} \tag{16}$$

$$F_f = F_f - \alpha \frac{\partial \Psi}{\partial F_f} \tag{17}$$

$$P_k = P_k - \alpha \frac{\partial \Psi}{\partial P_k} \tag{18}$$

Algorithm 1 SQPMF Model optimization.

Require: Matrix R, S and Q , learning rate α , regularization coefficient λ feature dimension W , maximum number of iterations $iters$;
Ensure: the predicted rating R_{ij}^z
 1: Initialize the feature matrices U, V, F, P ▷ to start SQPMF training
 2: **while** $i \neq iters$ **do** ▷ Calculate the derivative by (15) - (18)
 3: $U_i = U_i - \alpha \frac{\partial \Psi}{\partial U_i}$ ▷ Update feature matrix U
 4: $V_j = V_j - \alpha \frac{\partial \Psi}{\partial V_j}$ ▷ Update feature matrix V
 5: $F_f = F_f - \alpha \frac{\partial \Psi}{\partial F_f}$ ▷ Update feature matrix F
 6: $P_k = P_k - \alpha \frac{\partial \Psi}{\partial P_k}$ ▷ Update feature matrix P
 7: **end while**
 8: Compute $R_{ij}^z = U_i^T V_j + U_f^T V_z + V_j^T V_z$ ▷ by (37)

Where $\alpha \in [0,1]$ is the learning rate of the latent feature. Algorithm 1 describes the procedure of updating matrices U, V, F and P . The final prediction matrix R^z is obtained by

multiplying U and V . First, the rating matrix R , the social relationship matrix S and the POI transition matrix Q are input into the algorithm. The objective function is minimized step by step with SGD on U, V, F and P , assuming they have W -dimensional features. Finally, the predicted scores for the POIs can be obtained in R^z .

3.5 Derivation of the objective function

In this section, we prove that minimizing the previous objective function can maximize the probability of U, V, F, P that satisfy the matrix factorization of R, S, Q statistically. According to Bayes' Theorem:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)} \tag{19}$$

In our context, A is U, V, F, P , while B is R, S, Q , under the conditions of standard deviations of the Gaussian distribution $\sigma_R^2, \sigma_S^2, \sigma_Q^2, \sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2$.

Therefore, we need to maximize $P(U, V, F, P|R, S, Q, \sigma_R^2, \sigma_S^2, \sigma_Q^2, \sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2) \Theta$, which can be transformed as below according to the Bayes' Theorem:

$$\Theta = \frac{P(R, S, Q|U, V, F, P, \sigma_R^2, \sigma_S^2, \sigma_Q^2) \cdot P(U, V, F, P|\sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2)}{P(R, S, Q|\sigma_R^2, \sigma_S^2, \sigma_Q^2)} \tag{20}$$

Since R, S, Q are known, $P(R, S, Q|\sigma_R^2, \sigma_S^2, \sigma_Q^2)$ is a constant, we only need to maximize $P(R, S, Q|U, V, F, P, \sigma_R^2, \sigma_S^2, \sigma_Q^2) \cdot P(U, V, F, P|\sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2)$.

Likewise, we can transform $P(U, V, F, P|\sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2)$ into $P(U|\sigma_U^2) \cdot P(V|\sigma_V^2) \cdot P(F|\sigma_F^2) \cdot P(P|\sigma_P^2)$.

Now we just need to maximize

$$P(R|U, V, \sigma_R^2) \cdot P(S|U, F, \sigma_S^2) \cdot P(Q|V, P, \sigma_Q^2) \cdot P(U|\sigma_U^2) \cdot P(V|\sigma_V^2) \cdot P(F|\sigma_F^2) \cdot P(P|\sigma_P^2)$$

We assume that each observation value U_i, V_j, F_f, P_k are all independent and identically distributed, and the feature matrices U, V, F and P obey the Gaussian prior distribution with the mean value of 0 and the variance of $\sigma_U^2, \sigma_V^2, \sigma_F^2$ and σ_P^2 :

$$P(U|\sigma_U^2) = \prod_{i=1}^M N(U_i|0, \sigma_U^2)I \tag{21}$$

$$P(V|\sigma_V^2) = \prod_{j=1}^N N(V_j|0, \sigma_V^2)I \tag{22}$$

$$P(F|\sigma_F^2) = \prod_{f=1}^M N(F_f|0, \sigma_F^2)I \tag{23}$$

$$P(P|\sigma_p^2) = \prod_{k=1}^N N(P_k|0, \sigma_p^2)I \tag{24}$$

At the same time, it is assumed that the discrepancy between the real value and the predicted value of the users' ratings on the POIs obeys a Gaussian distribution with a mean value of 0 and a variance of σ_R^2 , then the conditional probability that the rating matrix R satisfies is shown in (25):

$$P(R|U, V, \sigma_R^2) = \prod_{i=1}^M \prod_{j=1}^N [N(R_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \tag{25}$$

Where I_{ij}^R is the indicator function, if the user i has rated the POI j , then its value equals 1; otherwise, it is 0. $g(x)$ maps the value of $U_i^T V_j$ to the interval of [0,1], where $g(x) = 1/(1 + e^{-x})$.

Then the probability distribution function satisfied by matrices S and Q is:

$$P(S|U, F, \sigma_S^2) = \prod_{i=1}^M \prod_{f=1}^M [N(S_{if}|g(U_i^T F_f), \sigma_S^2)]^{I_{if}^S} \tag{26}$$

$$P(Q|V, P, \sigma_Q^2) = \prod_{j=1}^N \prod_{k=1}^N [N(Q_{jk}|g(V_j^T P_k), \sigma_Q^2)]^{I_{jk}^Q} \tag{27}$$

Where I_{if}^S, I_{jk}^Q are indicator functions. $g(x)$ maps values to the interval [0,1], likewise, $g(x) = 1/(1 + e^{-x})$.

Therefore,

$$\begin{aligned} &P(R|U, V, \sigma_R^2) \cdot P(S|U, F, \sigma_S^2) \cdot P(Q|V, P, \sigma_Q^2) \cdot P(U|\sigma_U^2) \cdot P(V|\sigma_V^2) \\ &\cdot P(F|\sigma_F^2) \cdot P(P|\sigma_P^2) = \\ &\prod_{i=1}^M \prod_{j=1}^N [N(R_{ij}|g(U_i^T V_j), \sigma_R^2)]^{I_{ij}^R} \cdot \prod_{i=1}^M \prod_{f=1}^M [N(S_{if}|g(U_i^T F_f), \sigma_S^2)]^{I_{if}^S} \cdot \\ &\prod_{j=1}^N \prod_{k=1}^N [N(Q_{jk}|g(V_j^T P_k), \sigma_Q^2)]^{I_{jk}^Q} \cdot \prod_{i=1}^M N(U_i|0, \sigma_U^2 I) \\ &\cdot \prod_{j=1}^N N(V_j|0, \sigma_V^2 I) \cdot \\ &\prod_{i=1}^M N(F_i|0, \sigma_F^2 I) \cdot \prod_{j=1}^N N(P_j|0, \sigma_P^2 I) \end{aligned} \tag{28}$$

Take logarithm of each value in combination with probability distribution function of multivariate Gaussian distribution:

$$\ln N(U_i|0, \sigma_U^2 I) = \ln \left(-\frac{1}{2\pi \frac{W}{2} |\sigma_U^2 I|^{\frac{1}{2}}} \right) - \frac{U_i^T U_i}{2\sigma_U^2}$$

$$\begin{aligned} &= -\ln(|\sigma_U^2 I|^{\frac{1}{2}}) - \frac{U_i^T U_i}{2\sigma_U^2} + C_u \\ &= -\frac{1}{2} \ln(\sigma_U^2 W) - \frac{U_i^T U_i}{2\sigma_U^2} + C_u \\ &= -\frac{W}{2} \ln(\sigma_U^2) - \frac{U_i^T U_i}{2\sigma_U^2} + C_u \end{aligned} \tag{29}$$

Similarly, we can get:

$$\ln N(V_j|0, \sigma_V^2 I) = -\frac{W}{2} \ln(\sigma_V^2) - \frac{V_j^T V_j}{2\sigma_V^2} + C_V \tag{30}$$

$$\ln N(F_f|0, \sigma_F^2 I) = -\frac{W}{2} \ln(\sigma_F^2) - \frac{F_f^T F_f}{2\sigma_F^2} + C_F \tag{31}$$

$$\ln N(P_k|0, \sigma_P^2 I) = -\frac{W}{2} \ln(\sigma_P^2) - \frac{P_k^T P_k}{2\sigma_P^2} + C_P \tag{32}$$

$$\begin{aligned} &\ln N(R_{ij}|g(U_i^T V_j), \sigma_R^2) \\ &= -\frac{1}{2} \ln(\sigma_R^2) - \frac{(R_{ij} - g(U_i^T V_j))^T (R_{ij} - g(U_i^T V_j))}{2\sigma_R^2} \\ &+ C_R \end{aligned} \tag{33}$$

$$\begin{aligned} &\ln N(S_{if}|g(U_i^T F_f), \sigma_S^2) \\ &= -\frac{1}{2} \ln(\sigma_S^2) - \frac{(S_{if} - g(U_i^T F_f))^T (S_{if} - g(U_i^T F_f))}{2\sigma_S^2} \\ &+ C_S \end{aligned} \tag{34}$$

$$\begin{aligned} &\ln N(Q_{jk}|g(V_j^T P_k), \sigma_Q^2) \\ &= -\frac{1}{2} \ln(\sigma_Q^2) - \frac{(Q_{jk} - g(V_j^T P_k))^T (Q_{jk} - g(V_j^T P_k))}{2\sigma_Q^2} \\ &+ C_Q \end{aligned} \tag{35}$$

Substituting (29) - (35) into (28), the logarithm of the posterior distribution is shown in (36).

$$\begin{aligned} &\ln P(U, V, F, P|R, S, Q, \sigma_R^2, \sigma_S^2, \sigma_Q^2, \sigma_U^2, \sigma_V^2, \sigma_F^2, \sigma_P^2) \propto \\ &-\frac{1}{2\sigma_R^2} \sum_{i=1}^M \sum_{j=1}^N I_{ij} ((R_{ij} - g(U_i^T V_j))^2) \\ &-\frac{1}{2\sigma_S^2} \sum_{i=1}^M \sum_{f=1}^M I_{if} ((S_{if} - g(U_i^T F_f))^2) \\ &-\frac{1}{2\sigma_Q^2} \sum_{j=1}^N \sum_{k=1}^N I_{jk} ((Q_{jk} - g(V_j^T P_k))^2) \end{aligned}$$

$$\begin{aligned}
 & -\frac{1}{2\sigma_U^2} \sum_{i=1}^M U_i^T U_i - \frac{1}{2\sigma_V^2} \sum_{j=1}^N V_j^T V_j \\
 & -\frac{1}{2\sigma_F^2} \sum_{i=1}^M F_f^T F_f - \frac{1}{2\sigma_P^2} \sum_{j=1}^N P_k^T P_k \\
 & -\frac{1}{2} \left(\left(\sum_{i=1}^M \sum_{j=1}^N I_{ij}^R \right) \ln \sigma_R^2 + \left(\sum_{i=1}^M \sum_{f=1}^M I_{if}^S \right) \ln \sigma_S^2 \right. \\
 & \quad \left. + \left(\sum_{j=1}^N \sum_{k=1}^N I_{jk}^Q \right) \ln \sigma_Q^2 \right) \\
 & -\frac{1}{2} (MW \ln \sigma_U^2 + NW \ln \sigma_V^2 + MW \ln \sigma_F^2 \\
 & \quad + NW \ln \sigma_P^2) + C
 \end{aligned} \tag{36}$$

where C is a constant, independent of the parameters. Since the observed noise variance and prior variance are fixed, maximizing the log posterior of the latent features is equivalent to minimizing the sum of squared errors of the objective function, as shown in (10), except we add some regulators in the objective function.

3.6 Ranking

Finally, by continuously optimizing the objective function Ψ , we obtained the ideal matrices U, V, F, P . As can be seen from the previous description, in our proposed objective function Ψ , we consider the transition relationship between users' social relationships and POIs. Therefore, we can use the trained potential matrix to calculate a user's score for the next POI. Assuming that the current user i is at location j , the predicted score for the user to visit the next POI z , denoted as R_{ij}^z , can be described in (37):

$$R_{ij}^z = U_i^T V_z + U_f^T V_z + V_j^T V_z \tag{37}$$

Where, $U_i^T V_z$ represents the user's personal score for the next POI z , $U_f^T V_z$ represents the user's friend's personal score for the next POI z , and $V_j^T V_z$ represents the transition score of the current POI j , to the next POI z .

3.7 Complexity analysis

In our SQPMF method, the time complexity and space complexity are mainly based on the objective function Ψ to calculate. In addition, during the optimization process of SQPMF, when updating the user feature matrix U , POI feature matrix V , social feature matrix of users F , and relevance feature matrix P , the loss function and its derivatives need to be calculated, which is closely related to the dimension of

the feature matrix W . Therefore, the size of W determines the computational efficiency of SQPMF.

We use $O(\rho_R)$, $O(\rho_S)$ and $O(\rho_Q)$ to represent the number of non-zero elements in the rating matrix R , user social matrix S , and POI transition matrix Q , respectively. W is the dimension of the feature matrix. Therefore, The time complexity of $\frac{\partial \Psi}{\partial U_i}, \frac{\partial \Psi}{\partial V_j}, \frac{\partial \Psi}{\partial F_f}, \frac{\partial \Psi}{\partial P_k}$, is $O(\rho_R W + \rho_S W)$, $O(\rho_R W + \rho_Q W)$, $O(\rho_S W)$, $O(\rho_Q W)$. Thus, the time complexity of our objective function Ψ is $O(\rho_R W + \rho_S W + \rho_Q W)$.

In our algorithm, the sparse matrices R, S and Q are represented in a Compressed Sparse Row (CSR) format that only stores nonzero elements in a sparse matrix [46]. Therefore, for space complexity, the space occupied by the rating matrix R is $O(\rho_R)$, the space occupied by the user social matrix S is $O(\rho_S)$, the space occupied by the POI transition matrix Q is $O(\rho_Q)$, the space occupied by the feature vectors of users and POIs is $O(\rho_R W)$, the space occupied by the feature vectors of users and users' friends is $O(\rho_S W)$, and the space occupied by the feature vectors of POIs and transition POIs is $O(\rho_Q W)$. Thus, the space complexity is $O(\rho_R W + \rho_S W + \rho_Q W + \rho_R + \rho_S + \rho_Q)$, which can be simplified to $O(\rho_R W + \rho_S W + \rho_Q W)$.

In summary, the time complexity and space complexity of SQPMF are both linear, and can be scaled to large-scale datasets in practical applications.

4 Result and discussion

4.1 Datasets

We evaluate the performance of SQPMF using three public datasets, Gowalla¹, Foursquare² and Brightkite³, which are publicly available (see the links at the footnote). In this experiment, the data in Gowalla, Foursquare, and Brightkite are all real-world, public datasets. Among them, Gowalla was founded in 2009 and allows users to check in and share information through mobile devices. It is a social network platform that can provide users with social services. The Gowalla dataset includes 196,591 users, 1,280,969 POIs, 6,442,890 checkins, and 950,327 links, collected from February 2009 to October 2010. Foursquare is also a globally renowned location-based large-scale social service website launched in 2009. It has become the most popular social network, with up to 10 million user sets and 3.34 million daily views. The Foursquare dataset includes 114,324 users, 3,820,891 POIs, 22,809,624 checkins, and 607,333 links, collected from April 2012 to January 2014. Brightkite is a

¹ <http://snap.stanford.edu/data/loc-gowalla.html>

² <http://www.foursquare.com>

³ <https://snap.stanford.edu/data/loc-Brightkite.html>

Table 3 Dataset information

Dataset	Users	POIs	Check-ins	Links	Sparseness
Gowalla	9617	585	95956	269812	98.29%
Foursquare	32158	2409	423675	312397	99.45%
Brightkite	50686	48367	3228563	388180	99.87%

location-based social networking website created in 2007, allowing users to freely share locations within the community. The Brightkite dataset contains 58,228 users, 4,491,143 checkins, 428,156 links, and was collected from April 2008 to October 2010. During the experiment, we select 80% of the dataset as the training dataset, and the remaining 20% as the test set.

In order to reduce noise in the datasets, we preprocessed three datasets. For each dataset, we removed users and POIs with less than 10 successive check-ins because users with few check-ins are inactive. Similarly, POIs with too few check-ins are not attractive. For the Gowalla dataset, after preprocessing, the number of users is 9617, the number of POIs is 585, the number of check-ins is 95,956, the number of user relationships is 269,812, and the sparseness of the user-POIs check-in matrix is 98.29%. For the Foursquare dataset, after data preprocessing, the number of users is 32,158, the number of POIs is 2,409, the number of check-ins is 423,675, the number of user relationships is 312,397, and the sparseness of the user-POI check-in matrix is 99.45%. For the Brightkite dataset, after data preprocessing, the number of users is 50,686, the number of POIs is 48,367, the number of check-ins is 3,228,563, the number of user relationships is 388,180, and the sparseness of the user-POI check-in matrix is 99.87%. The details of the datasets are shown in Table 3.

4.2 Evaluation metrics

We use five metrics, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Precision, Recall and F1-score for performance evaluation. Among them, MAE and RMSE are the most commonly used metrics to measure the error of recommendation results in recommender systems. The smaller the values of MAE and RMSE, the smaller the error between the real value and the predicted value, and the better performance of an algorithm. Besides, Precision and Recall are used to evaluate the accuracy of the recommendation results. The higher the values of Precision and Recall, the higher the accuracy of the recommendation and the better performance of algorithm. The F1-score is the weighted average of Precision and Recall, with a range of 0 to 1. Generally speaking, the closer the F1-score is to 1, the better the performance of the recommendation model.

The definitions of the metrics are as follows:

MAE:

$$MAE = \frac{\sum_{i,j} |R_{ij} - R'_{ij}|}{N} \quad (38)$$

RMSE:

$$RMSE = \sqrt{\frac{\sum_{i,j} (|R_{ij} - R'_{ij}|)^2}{N}} \quad (39)$$

Precision:

$$Precision@K = \frac{\sum_U |R(U) \cap T(U)|}{R(U)} \quad (40)$$

Recall:

$$Recall@K = \frac{\sum_U |R(U) \cap T(U)|}{T(U)} \quad (41)$$

F1-score:

$$F1 - score = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (42)$$

Among them, R_{ij} is the rating score of user i on POI j , R'_{ij} refers to the predicted rating score of user i on POI j , N refers to the total number of scores in the test set, and $R(U)$ is the $Top - k$ list, $T(U)$ is the number of POIs actually visited by the user.

4.3 Related models for comparison

We selected several popular existing models in the field of PMF and POI recommendation for comparison with SQPMF. Among them, PMF is a typical PMF model. Sorec and TrustSVD consider the users' social trust relationships in the recommendation. The GeoMF and PRME are both typical models in the field of POI recommendation. The SSTPMF and RTPM are the latest method in the field of successive POI recommendation. A brief introduction of these models is as follows:

- (1) PMF [26]: this method applies probability-related knowledge in MF, decomposes user-POI matrix into user feature matrix and POI feature matrix, and assumes that both user feature matrix and POI feature matrix obey Gaussian distribution with mean 0.
- (2) Sorec [34]: this method is based on the PMF model, and proposes a new social recommendation framework, which integrates the user's trust relationship matrix.
- (3) TrustSVD [1]: this method incorporates the effect of trust into SVD++, where the strength of trust is set equally for all users.

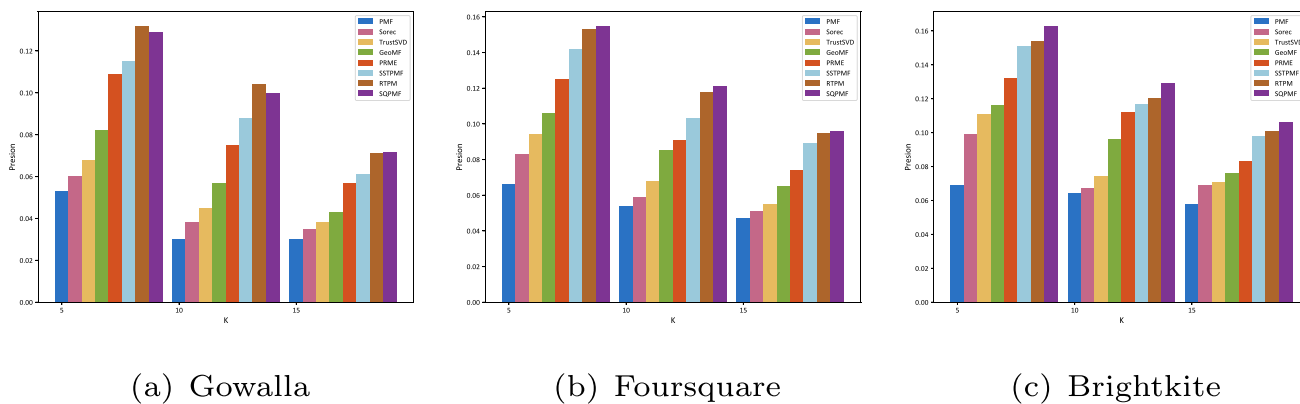


Fig. 3 Precision comparison with different top-k

- (4) GeoMF [37]: this method uses the MF technique to make recommendations based on the geographical model of the user’s activity area and the influence area of POIs.
- (5) PRME [29]: this method combines sequence information with personal preferences by mapping each POI into a low-dimensional Euclidean potential space of the target, and then uses a metric embedding algorithm to efficiently compute POI transition in a Markov chain model.
- (6) SSTPMF [27]: this method is based on the PMF model, and considering POI similarity and user similarity, this model proposes a multivariable inference approach for POI recommendation using latent similarity factors.
- (7) RTPM [39]: this method proposes a real-time preference mining model based on LSTM, mining users’ real-time preferences from long-term and short-term preferences to recommend the next POI with time constraints.

4.4 Comparison with related models

In this section, we compare SQPMF with PMF, Sorec, TrustSVD, GeoMF, PRME, SSTPMF and RTPM in terms of recommendation Precision, Recall and F1-score. The value of the abscissa K is the number of POIs recommended to the user.

According to the experimental results in Figs. 3, 4 and 5, we took the average of the experimental results from three datasets, our SQPMF improves Precision, Recall, and F1-score by at least more than 6.1%, 5.8% and 5.7%, respectively compared with the second best RTPM.

Among the compared methods, PMF is a direct application of the PMF model. However, the recommendation of PMF is the worst, because PMF only considers the users’ preferences for POIs without using other contextual factors. Sorec and TrustSVD have integrated the users’ social relationships in

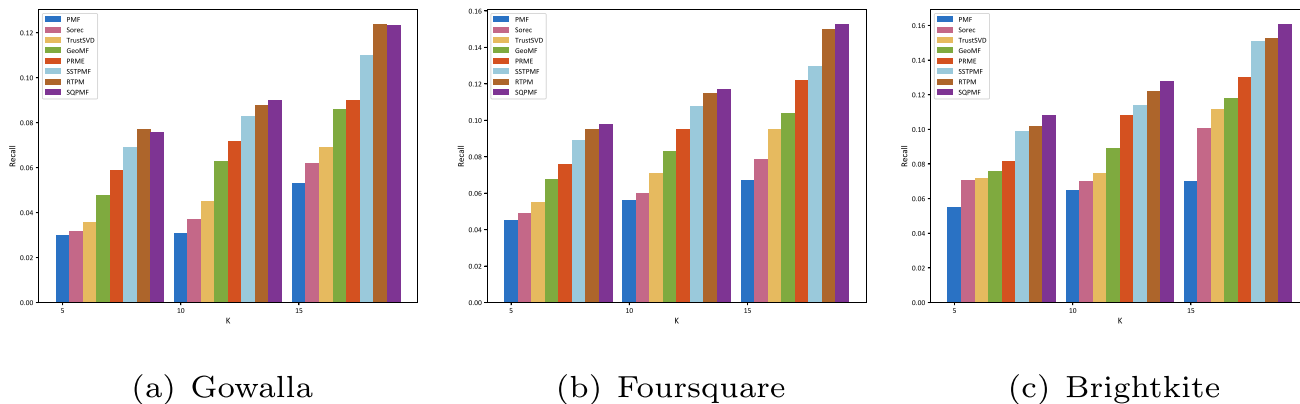


Fig. 4 Recall comparison with different top-k

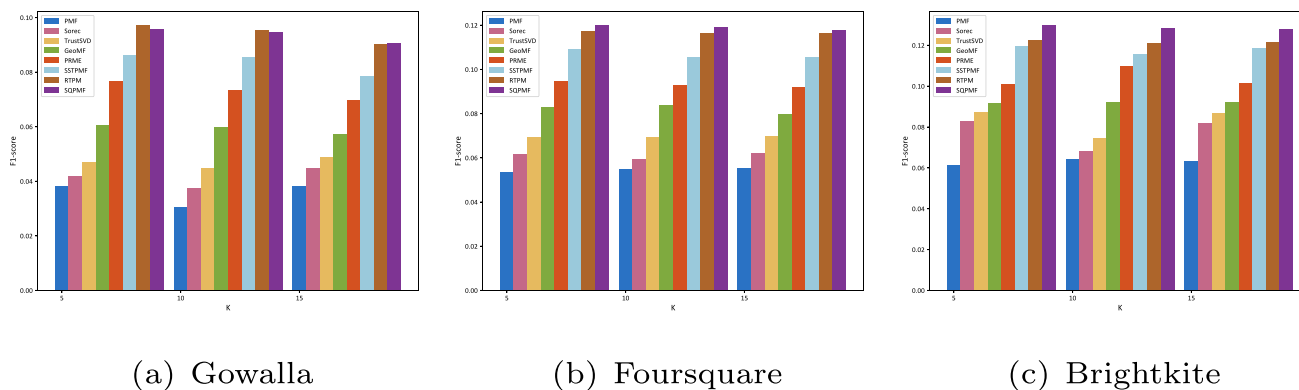


Fig. 5 F1-score comparison with different top-k

the recommendation, so they clearly improve the recommendation performance. PRME further integrates the check-in sequence information and personal preference information, so it further improves the recommendation performance. GeoMF and PRME have used POI transition relationships, which show the impact of POI transition relationships on recommendation results.

However, the above models only consider partial relationships between users or between POIs, and they do not take advantage of both relationships simultaneously. Therefore, our experimental results in Fig. 3 show that they do not perform as well as SSTPMF and our SQPMF, which use both relationships.

SSTPMF considers the relevant impacts of users and POIs, and therefore has improved the accuracy of POI recommendations. However, although SSTPMF considers the relationships between users and POIs simultaneously, it is not accurate enough in calculating the similarity between users and POIs. For example, SSTPMF only uses Jackard similarity to calculate relationships between users, but it does not consider the users' habit similarity and sequence similarity as we do in SQPMF. Similarly, SSTPMF also has the problem of inaccurate mapping of POI category when calculating POI similarity.

For large datasets such as Foursquare and Brightkite that contain more check-in information, SQPMF performs 8.8% and 10.6% better than RTPM, respectively. This is because RTPM, while considering the periodicity of user behavior patterns when constructing long-term preferences, only utilizes the influence of distance between POIs for modeling, without considering other POI transition relationships. When constructing users' short-term preferences, RTPM does not consider the personal attributes and current location of users who have a strong impact on the next POI, but only considers the influence of the public, which greatly affects recommendation performance. Therefore, when there is a large amount of check-in data in the dataset, there is a large amount

of implicit information between users and POIs, and the performance of RTPM is average. However, SQPMF takes full advantage of the social relationships of users and the sequential transition relationships between successive POIs, demonstrating better performance.

It is worth noting, in the Gowalla dataset, RTPM performs slightly better than SQPMF. This is because, in order to obtain ratings with minimal errors, SQPMF needs a large number of social relationships between users and POIs for better training and prediction. However, the Gowalla dataset has lower average number of check-in times per user, so there is not enough check-in data in the Gowalla dataset for SQPMF to mine the relationships between users and POIs and to exhibit the best performance. This suggests that SQPMF is better suited to large sparse datasets with many active users, which are consistent with reality.

Our SQPMF method better integrates the users' personal preferences, the users' social relationships and the POI transition relationships between successive POIs in the recommendation system. From the results, we can see that the transition of successive POIs also has a significant effect

Table 4 Parameter description and value

Parameter	Value	Description
W	20	Feature dimension
λ	0.5	Regularization coefficient
η	0.6	The weight of trust degree
μ	0.2	The weight of habit similarity
λ_S	0.2	The regularization coefficient of the social relationship matrix
β	0.7	The weight of transition frequency score
γ	0.2	The weight of distance score
λ_Q	0.1	The regularization coefficient of the POI transition matrix
α	0.01	Learning rate

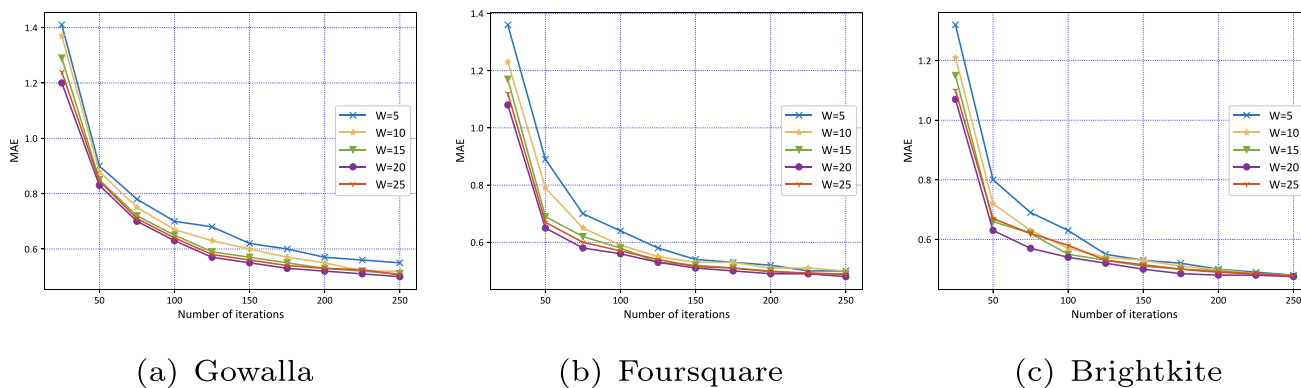


Fig. 6 MAE of SQPMF with different iterations

on the recommendation results. Therefore, SQPMF achieves the highest Precision, Recall and F1-score values on three datasets.

It is worth noting that, for all methods, when K becomes larger, the precision decreases, but the recall increases. This is because, in real life, when the number of POI recommended to users increases, the users are more willing to visit new POI rather than the visited places.

4.5 Impact of parameters on SQPMF

In our previous experiments, the specified feature dimension W in the SQPMF model is set to 20. The regularization coefficient λ is set to 0.5. The weight coefficient η of the social score is 0.6, μ is 0.2, the regularization coefficient λ_S of the social relationship matrix is set to 0.2, the weight coefficient β of the transition score is 0.7, γ is 0.2, and the regularization coefficient λ_Q of the POI transition matrix is set to 0.1. The regularization coefficients of the features of users, features of POIs, features of social user, and features of relevance POIs are all set to 0.001, and the learning rate α is set to 0.01.

These parameters are chosen as they give the best results for SQPMF. The parameter values are shown in Table 4.

For the dimension of the specified feature W , in order to find the optimal value, we select the values of 5, 10, 15, 20 and 25 for W . Figures 6 and 7 give the results of MAE and RMSE on two datasets in relation to the number of latent features W used in SQPMF.

According to these figures, when W increases, SQPMF can achieve better performance. We can see that the different values of W can make MAE differ by 2%-4%. This shows W has a significant impact. Note that when W is set to 20, our experimental results achieve the best performance. According to the complexity mentioned above, we can know that the time complexity of the system is closely related to the value of W . If W is too larger, say 25, it does not bring better recommendation effect, but significantly increases the computational cost. Therefore, we choose 20 for W according to our experiments.

In the calculation of social scores, η represents the weight of the trust degree, μ represents the weight of the habit similarity, and $(1 - \eta - \mu)$ represents the weight of the sequence similarity. According to our experiment, we set η to be in the

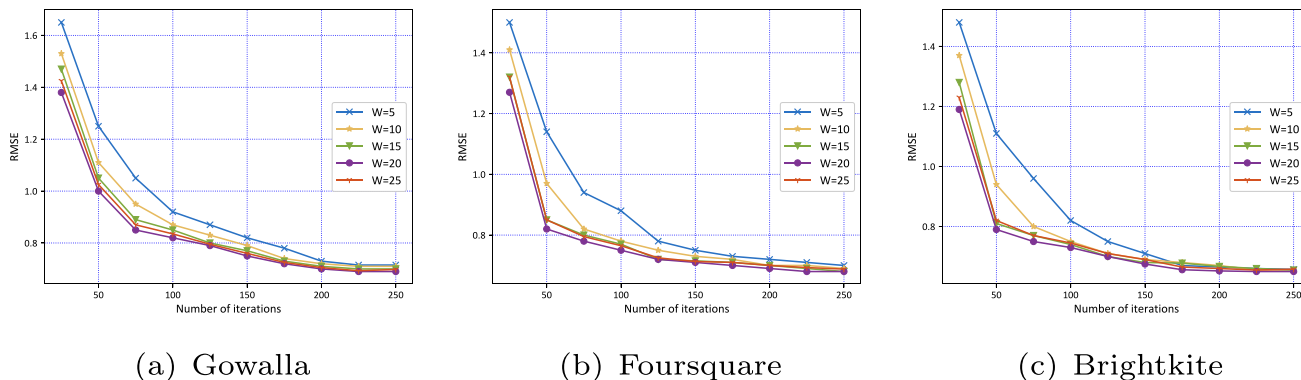


Fig. 7 RMSE of SQPMF with different iterations

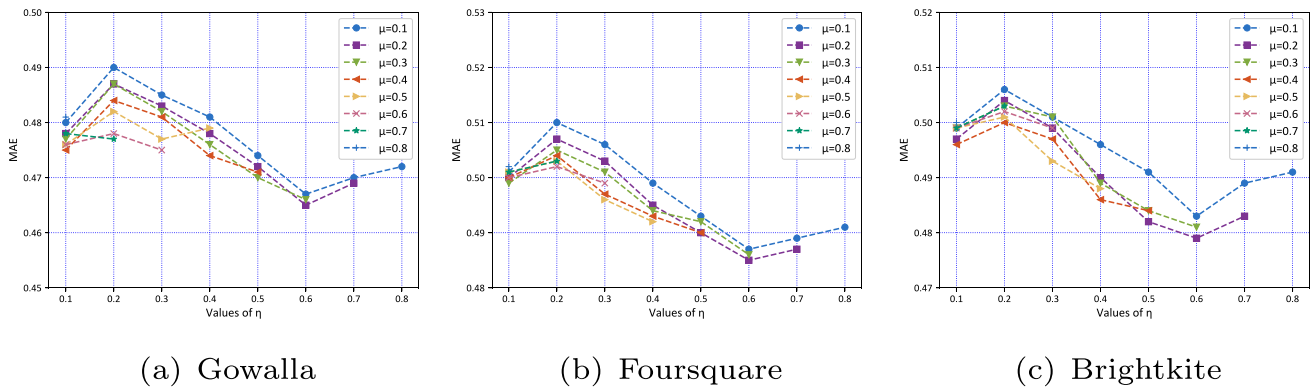


Fig. 8 MAE of SQPMF with different η and μ

range of [0.1,0.8], to test the impact of μ on MAE and RMSE. Similarly, when calculating the transition score, β represents the weight of the transition frequency score, γ represents the weight of the geographical distance score, and $(1 - \beta - \gamma)$ represents the weight of the popularity score of destination. During the experiment, we set the value of β to be in the range of [0.1,0.8], change the value of γ . Finally, according to the experimental results, we determine the optimal values of the parameters.

From Figs. 8 and 9 we can see that when η increases to 0.6 and μ decreases to 0.2, the recommendation effect is the best at this time. Similarly, from Figs. 10 and 11, we can find that when β is 0.7 and γ is 0.2, the MAE and RMSE can reach the best values.

These results show that in the social relationship matrix, the trust degree is the most influential in the user’s social relationship, and the habit similarity and the sequence similarity have similar effects. For the POI transition matrix, the most influential factor is transition frequency, followed by geographic distance.

The optimization of the objective function involves the regularization coefficient λ_S of the social relationship matrix and the regularization coefficient λ_Q of the POI transition matrix. When λ_S is set to 0, the system does not consider the user’s social influence factor in recommendation, but only considers the user’s personal preferences and the POI transition relationships. Similarly, when λ_Q is set to 0, the system does not consider POI transition relationships in recommendation, but only considers the user’s personal preferences and the user’s social relationships.

From Figs. 12 and 13, we can see that, when λ_S and λ_Q changes constantly, MAE and RMSE change accordingly. When λ_S is set to 0.2, and λ_Q is set to 0.1, MAE and RMSE reach to the lowest values, which indicate the performance is the best at these points.

The learning rate α of gradient descent also has a significant impact on the performance (MAE and RMSE). From Fig. 14 we can see that, when α is 0.01, MAE and RMSE reach to the smallest values. This is because, if α is too small, there need more iterations to make the objective function

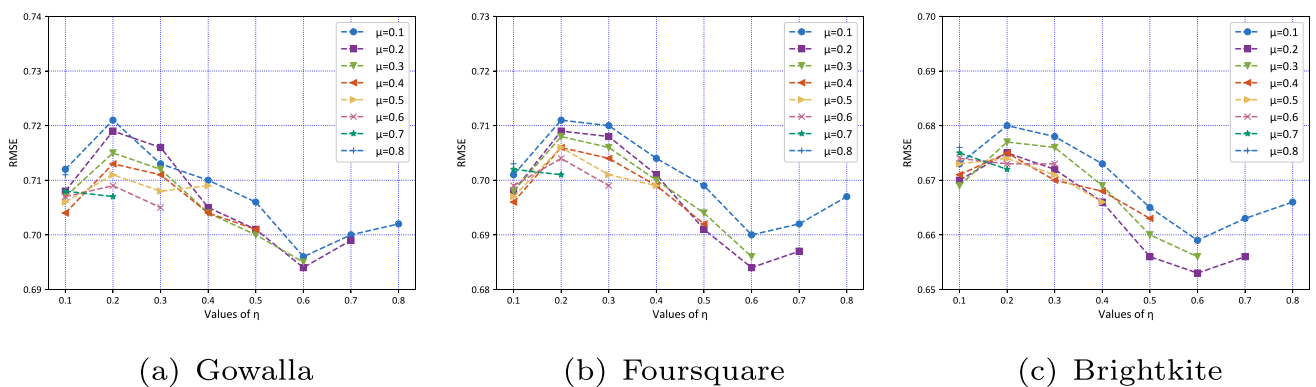


Fig. 9 RMSE of SQPMF with different η and μ

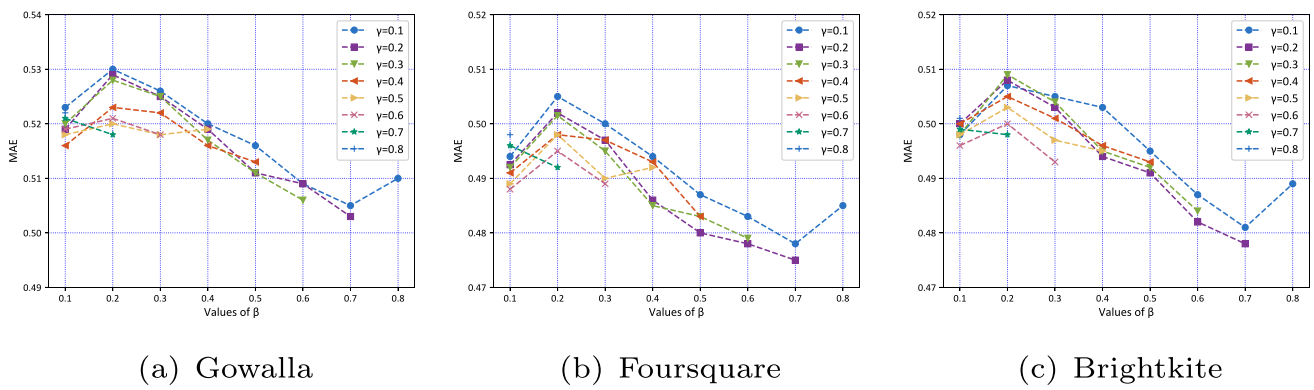


Fig. 10 MAE of SQPMF with different β and γ

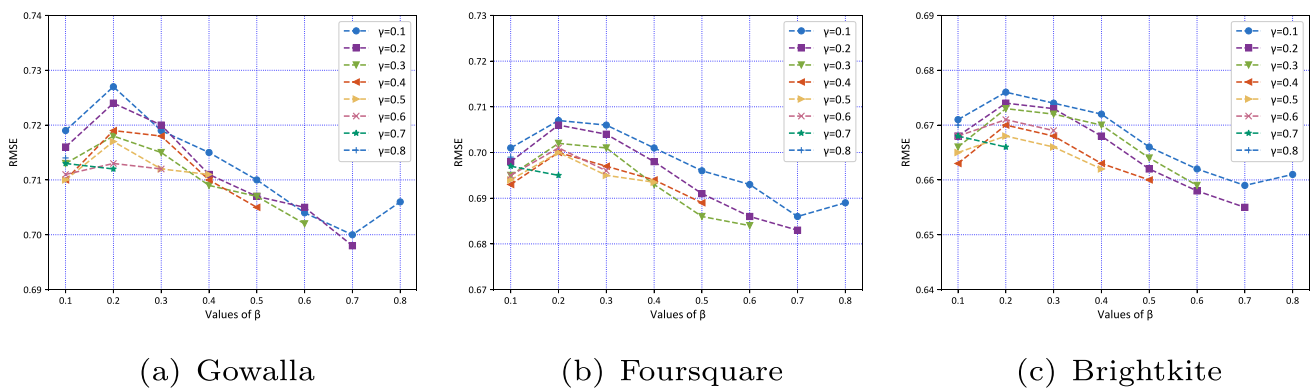


Fig. 11 RMSE of SQPMF with different β and γ

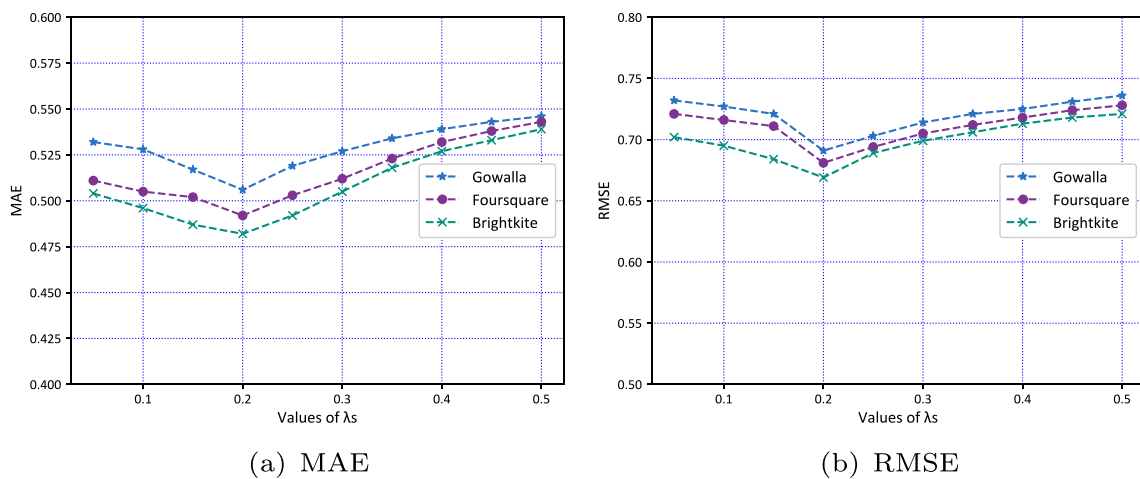


Fig. 12 MAE&RMSE of SQPMF with different λ_S

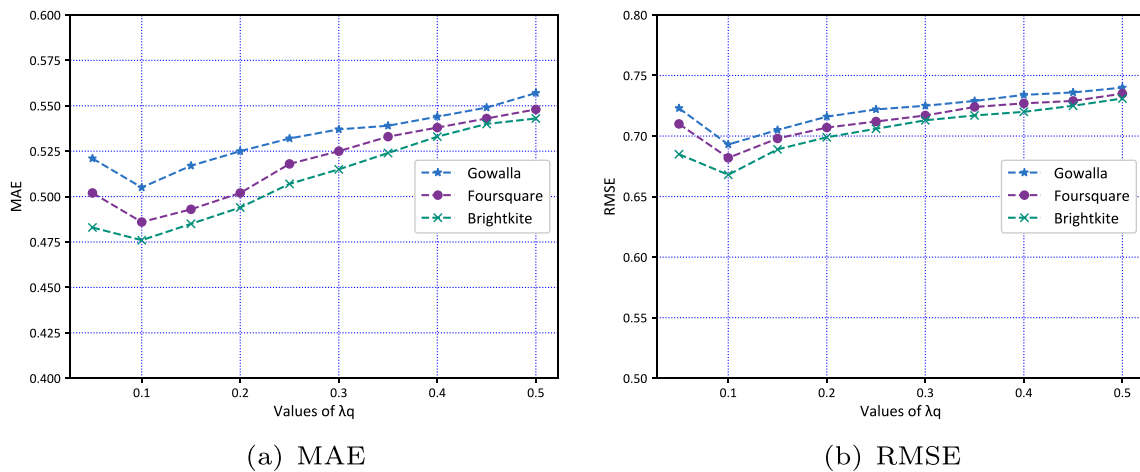


Fig. 13 MAE&RMSE of SQPMF with different λ_Q

converge. If α is too large, the objective function will fluctuate greatly during the training process, which may miss the minimum value or is unable to converge.

5 Conclusion

In this paper, we propose a new method, called SQPMF, that fuses the users' personal preferences, the users' social relationships and the transition relationships between successive POIs in POI recommendation. By considering these contextual information, SQPMF significantly improves the accuracy of recommendation results. Specifically, we deconstruct the users' social relationships into several important factors and build a social relationship matrix that integrates trust relationships, user habits, and user sequence similarity. We also deconstruct the transition relationships between

successive POIs into factors such as transition frequency, the geographic distance, and the popularity of destination, and build a POI transition matrix that reflects the transition relationships of successive POI sequences. By embedding the social relationship matrix and the POI transition matrix into the solution, SQPMF can effectively predict and recommend the user's next POI based on three real datasets. Our experimental results show that SQPMF achieves the highest recommendation accuracy for sparse datasets among the state-of-the-art methods.

SQPMF relies on accurate information about users' trust relationships and POI transition relationships. In the field of successive POI recommendation, there is still a lot of implicit information that we could further exploit to improve the prediction accuracy of SQPMF. In future, we expect to be able to mine more effective implicit information, such as spatio-temporal information, user dynamic preferences, etc.,

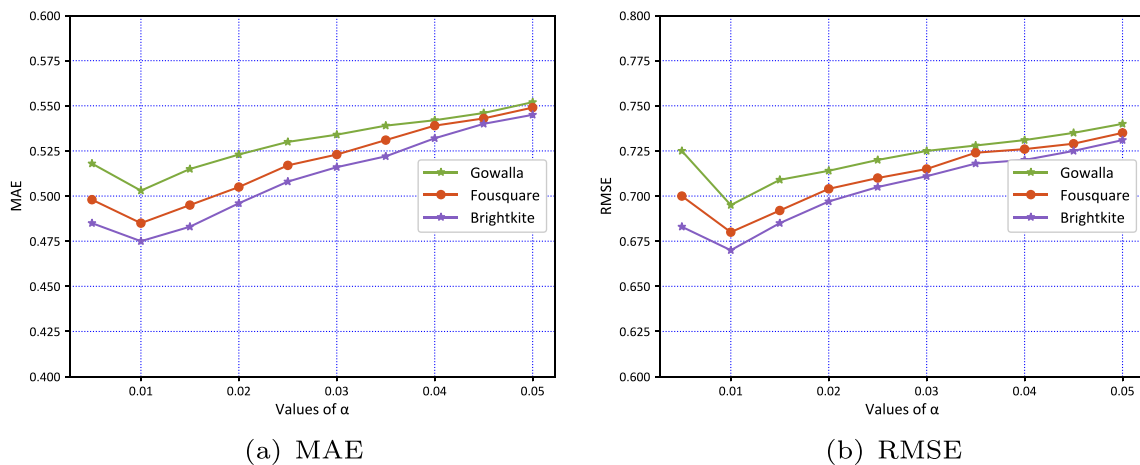


Fig. 14 MAE&RMSE of SQPMF with different α

to further improve the accuracy of recommendations. In addition, the issue of cold start is also a highly concerned issue in the field of successive POI recommendation. SQPMF has not yet showed good performance in solving cold start problems, which will be one of our future efforts. Also we will focus on the privacy protection problem in the POI recommendation system and hope to better trade off between privacy and recommendation.

Acknowledgements This work is supported by the National Nature Science Foundation of China 61370198 and 61300187, and in part by the Liaoning Provincial Natural Science Foundation of China under Grant 2019- MS-028.

Author Contributions Jie Wang: Conceptualization, Methodology, Software, Investigation, Writing-Original Draft. Zhiyi Huang: Formal analysis, Supervision, Writing-Review & Editing. Zhaobin Liu: Formal analysis, Methodology, Writing-Review & Editing.

Data availability and access Data will be made available from the corresponding author on reasonable request.

Declarations

Informed consent The datasets generated or analyzed during this study are available in <http://snap.stanford.edu/data/loc-gowalla.html>, <http://www.foursquare.com>, <https://snap.stanford.edu/data/loc-Brightkite.html>

Competing Interests All authors have no relevant financial or non-financial interests to disclose.

References

- Guo G, Zhang J, Yorke-Smith N (2015) Trustsvd: collaborative filtering with both the explicit and implicit influence of user trust and of item ratings. In: Proceedings of the AAAI conference on artificial intelligence, vol 29
- Wang H, Shen H, Ouyang W, Cheng X (2018) Exploiting point-specific geographical influence for point-of-interest recommendation. In: IJCAI, pp 3877–3883
- Lim KH, Chan J, Leckie C, Karunasekera S (2018) Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. *Knowl Inf Syst* 54:375–406
- Guo L, Wen Y, Liu F (2019) Location perspective-based neighborhood-aware poi recommendation in location-based social networks. *Soft Comput* 23(22):11935–11945
- Islam MA, Mohammad MM, Das SSS, Ali ME (2022) A survey on deep learning based point-of-interest (poi) recommendations. *Neurocomputing* 472:306–325
- Liu S, Yang L, Zheng W, Xiao Y, Liu L (2022) An ensemble learning model for preference-geographical aware point-of interest recommendation. *Appl Intell* 52(12):13763–13780
- Li X, Hu R, Wang Z (2022) Next-point-of-interest recommendation based on joint mining of regularity and randomness. *Knowl-Based Syst* 258:110052
- Zhou Y, Yang G, Yan B, Cai Y, Zhu Z (2022) Point-of-interest recommendation model considering strength of user relationship for location-based social networks. *Expert Syst Appl* 199:117147
- Qi L, Liu Y, Zhang Y, Xu X, Bilal M, Song H (2022) Privacy-aware point-of-interest category recommendation in internet of things. *IEEE Internet Things J* 9(21):21398–21408
- Sun Q, Chiclana F, Wu J, Liu Y, Liang C, Herrera-Viedma E (2023) A weight penalty mechanism for non-cooperative behavior in large-scale group decision making with unbalanced linguistic term sets. *IEEE Transactions on fuzzy systems*
- Naserian E, Wang X, Dahal KP, Alcaraz-Calero JM, Gao H (2021) A partition-based partial personalized model for points-of-interest recommendations. *IEEE Transactions on computational social systems* 8(5):1223–1237
- Cao G, Cui S, Joe I (2023) Improving the spatial-temporal aware attention network with dynamic trajectory graph learning for next point-of-interest recommendation. *Information Processing & Management* 60(3):103335
- Huang L, Ma Y, Wang S, Liu Y (2019) An attention-based spatiotemporal lstm network for next poi recommendation. *IEEE Trans Serv Comput* 14(6):1585–1597
- Lu Y-S, Shih W-Y, Gau H-Y, Chung K-C, Huang J-L (2019) On successive point-of-interest recommendation. *World Wide Web* 22:1151–1173
- Baral R, Iyengar SS, Zhu X, Li T, Sniatala P (2019) Hirecs: a hierarchical contextual location recommendation system. *IEEE Transactions on Computational Social Systems* 6(5):1020–1037
- Rahmani HA, Deldjoo Y, Di Noia T (2022) The role of context fusion on accuracy, beyond-accuracy, and fairness of point-of-interest recommendation systems. *Expert Syst Appl* 205:117700
- Cai Z, Yuan G, Qiao S, Qu S, Zhang Y, Bing R (2022) Fg-cf: friends-aware graph collaborative filtering for poi recommendation. *Neurocomputing* 488:107–119
- Choudhary C, Singh I, Kumar M (2023) Sarwas: deep ensemble learning techniques for sentiment based recommendation system. *Expert Syst Appl* 216:119420
- Li J, Lu K, Huang Z, Shen HT (2019) On both cold-start and long-tail recommendation with social data. *IEEE Trans Knowl Data Eng* 33(1):194–208
- Vitoropoulou M, Tsitsekis K, Karyotis V, Papavassiliou S (2021) Cover: an information diffusion aware approach for efficient recommendations under user coverage constraints. *IEEE Transactions on Computational Social Systems* 8(4):894–905
- Qian X, Feng H, Zhao G, Mei T (2013) Personalized recommendation combining user interest and social circle. *IEEE Transactions Knowledge Data Engineering* 26(7):1763–1777
- Xing S, Liu F, Zhao X, Li T (2018) Points-of-interest recommendation based on convolution matrix factorization. *Appl Intell* 48:2458–2469
- Liu Y, Wu H, Rezaee K, Khosravi MR, Khalaf OI, Khan AA, Ramesh D, Qi L (2022) Interaction-enhanced and time-aware graph convolutional network for successive point-of-interest recommendation in traveling enterprises. *IEEE Transactions Industrial Informatics* 19(1):635–643
- Rahmani HA, Aliannejadi M, Ahmadian S, Baratchi M, Afsharchi M, Crestani F (2020) Lglmf: local geographical based logistic matrix factorization model for poi recommendation. In: Information retrieval technology: 15th Asia information retrieval societies conference, AIRS 2019, Hong Kong, China, November 7–9, 2019, Proceedings 15, Springer, pp 66–78
- Seyedhoseinzadeh K, Rahmani HA, Afsharchi M, Aliannejadi M (2022) Leveraging social influence based on users activity centers for point-of-interest recommendation. *Inform Process & Manag* 59(2):102858
- Mnih A, Salakhutdinov RR (2007) Probabilistic matrix factorization. *Adv Neural Inform Process Syst* 20
- Davtalab M, Alesheikh AA (2021) A poi recommendation approach integrating social spatio-temporal information into probabilistic matrix factorization. *Knowl Inf Syst* 63:65–85

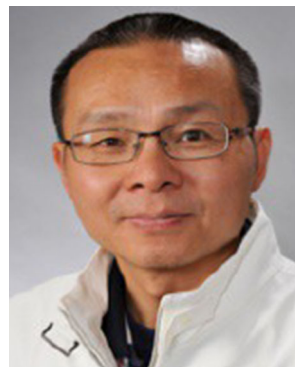
28. Wu L, Chen E, Liu Q, Xu L, Bao T, Zhang L (2012) Leveraging tagging for neighborhood-aware probabilistic matrix factorization. In: Proceedings of the 21st ACM international conference on information and knowledge management, pp 1854–1858
29. Feng S, Li X, Zeng Y, Cong G, Chee YM (2015) Personalized ranking metric embedding for next new poi recommendation. In: IJCAI'15 Proceedings of the 24th international conference on artificial intelligence, ACM, pp 2069–2075
30. Bao J, Zheng Y, Wilkie D, Mokbel M (2015) Recommendations in location-based social networks: a survey. *GeoInformatica* 19:525–565
31. Amirat H, Lagraa N, Fournier-Viger P, Ouinten Y, Kherfi ML, Guelouma Y (2023) Incremental tree-based successive poi recommendation in location-based social networks. *Appl Intell* 53(7):7562–7598
32. Chen Y-C, Thaipisutikul T, Shih TK (2020) A learning-based poi recommendation with spatiotemporal context awareness. *IEEE Transaction on cybernetics* 52(4):2453–2466
33. Xiong F, Shen W, Chen H, Pan S, Wang X, Yan Z (2019) Exploiting implicit influence from information propagation for social recommendation. *IEEE Transactions on Cybernetics* 50(10):4186–4199
34. Ma H, Yang H, Lyu MR, King I (2008) Sorec: social recommendation using probabilistic matrix factorization. In: Proceedings of the 17th ACM conference on information and knowledge management, pp 931–940
35. Chen W, Wan H, Guo S, Huang H, Zheng S, Li J, Lin S, Lin Y (2022) Building and exploiting spatial-temporal knowledge graph for next poi recommendation. *Knowl-Based Syst* 258:109951
36. Hu B, Ye Y, Zhong Y, Pan J, Hu M (2022) Transmkr: Translation-based knowledge graph enhanced multi-task point-of-interest recommendation. *Neurocomputing* 474:107–114
37. Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y (2014) Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining, pp 831–840
38. Liu J, Chen Y, Huang X, Li J, Min G (2023) Gnn-based long and short term preference modeling for next-location prediction. *Inf Sci* 629:1–14
39. Liu X, Yang Y, Xu Y, Yang F, Huang Q, Wang H (2022) Real-time poi recommendation via modeling long-and short-term user preferences. *Neurocomputing* 467:454–464
40. Zhang Z, Dong M, Ota K, Zhang Y, Ren Y (2021) Lbcf: a link-based collaborative filtering for overfitting problem in recommender system. *IEEE Transactions on Computational Social Systems* 8(6):1450–1464
41. Fang J, Meng X (2022) Urpi-gru: an approach of next poi recommendation based on user relationship and preference information. *Knowl-Based Syst* 256:109848
42. Zheng C, Tao D, Wang J, Cui L, Ruan W, Yu S (2020) Memory augmented hierarchical attention network for next point-of-interest recommendation. *IEEE Transactions on Computational Social Systems* 8(2):489–499
43. Wang H, Li P, Liu Y, Shao J (2021) Towards real-time demand-aware sequential poi recommendation. *Inf Sci* 547:482–497
44. Ji F, Wu J, Chiclana F, Wang S, Fujita H, Herrera-Viedma E (2023) The overlapping community driven feedback mechanism to support consensus in social network group decision making. *IEEE Transactions on fuzzy systems*
45. Aliannejadi M, Rafailidis D, Crestani F (2019) A joint two-phase time-sensitive regularized collaborative ranking model for point of interest recommendation. *IEEE Trans Knowl Data Eng* 32(6):1050–1063
46. Borstnik U, Vandevondevle J, Weber V, Hutter J (2014) Sparse matrix multiplication: the distributed block-compressed sparse row library. *Parallel Comput* 40(5–6), 47–58

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Jie Wang is currently working toward the M.E degree in the College of Information Science and Technology, Dalian Maritime University, China. She received the B.E. degree from Dalian Minzu University, China. Her currently research interests include machine learning, data mining and data privacy.



networks.

Zhiyi Huang received his PhD degree in Computer Science in 1992 from National University of Defense Technology, China. He is currently a Professor in the University of Otago, New Zealand. His research areas include artificial neural networks, parallel/distributed computing, signal processing on EEG data, machine learning algorithms, multi-core/manycore architectures, cluster computing, parallel programming environments, operating systems, and computer



Zhaobin Liu is a professor in the College of Information Science and Technology, Dalian Maritime University, China. He received the Ph.D degree in computer science from Huazhong University of Science and Technology, China, in 2004. His research interests include big data, machine learning, cloud computing and data privacy.